TOTAL
Editor: Călin Vlasie
Redactor: Ioan Es. Pop Tehnoredactare: Gabriela Chircea Coperta colecţiei: Ionuţ Broştianu Prepress: Marius Badea
Descrierea CIP a Bibliotecii Naţionale a României
KURZWEIL, RAY
Cum se construieşte o minte: secretul dezvăluit al gândirii umane / Ray Kurzweil; trad.: Bogdan Chircea.
— Piteşti: Paralela 45, 2013
ISBN 978 – 973 – 47 – 1651 – 7
I. Chircea, Bogdan (trad.)
004
Fotografia lui Rosalind Franklin de la pag. 22 este folosită cu acordul National Portrait Gallery Londra:
NPG X76929 Rosalind Elsie Franklin vby Elliott & Fry half-plate film negative, 11 June 1946 e National Portrait Gallery, London
Ray Kurzweil
How to create a mind. The secret of human thought revealed
Copyright © Ray Kurzweil, 2012
Published by arrangement with Viking, a member of Penguin Group (USA) Inc.
All rights reserved including the right of reproduction in whole or în part in any form.
Copyrighte Editura Paralela 45, 2013, pentru prezenta ediţie
Ray Kurzweil
CUM SE CONSTRUIEŞTE O MINTE
Secretul dezvăluit al gândirii umane
Traducere din limba engleză: *Bogdan Chircea*

PARALELA 45

Îi dedic această carte tânărului Kurzweii al României, Ionuţ *Budişteanu, laureat al Premiului Turing – Nobelul informaticii. Pentru geniul lui şi pentru speranţa pe care mi-o dă în afirmarea şi împlinirea posibilităţilor fără limite*

ale inteligenţei omului. Şi pentru că mi-a dat ocazia să văd pe viu ce înseamnă o minte cu adevărat strălucită.

Bogdan Chircea

Dedicaţie *Lui Leo Oscar Kurzweil. Intri într-o lume extraordinară.*

MULŢUMIRI

f

Vreau să-mi exprim gratitudinea faţă de soţia mea, Sonya, pentru răbdarea ei plină de iubire în faţa vicisitudinilor procesului creator;

Faţă de copiii mei, Ethan şi Amy, de nora mea, Rebecca, de sora mea, Enid, şi de nepotul meu, Leo, pentru iubire şi inspiraţie;

Faţă de mama mea, Hannah, pentru sprijinirea ideilor şi invenţiilor mele de la începuturi, care mi-au dat libertatea să experimentez de la o vârstă fragedă şi pentru că l-a ţinut pe tatăl meu în viaţă pe parcursul îndelungatei lui suferinţe;

Faţă de editorul meu de multă vreme de la Viking, Rick Kot, pentru spiritul lui de conducător, pentru călăuzirea permanentă şi plină de pătrundere şi pentru redactarea expertă;

Faţă de Loretta Barrett, agentul meu literar de douăzeci de ani, pentru îndrumarea abilă şi plină de entuziasm;

Faţă de Aaron Kleiner, vechiul meu partener de afaceri, pentru colaborarea plină de devotament din ultimii patruzeci de ani;

Faţă de Amara Angelica, pentru sprijinul devotat şi excepţional pe care mi l-a acordat în munca de documentare;

Faţă de Sarah Black, pentru intuiţia şi ideile extraordinare privitoare la documentare;

Faţă de Laksman Frank, pentru ilustraţiile de excepţie;

Faţă de Sarah Reed, pentru sprijinul entuziast privitor la organizare;

Faţă de Nanda Barker-Hook, pentru organizarea la nivel de expert a evenimentelor mele publice pe această temă, dar şi pe altele;

Faţă de Amy Kurzweil, pentru îndrumarea în meşteşugul scrisului;

Faţă de Cindy Mason, pentru sprijinul în documentare, pentru ideile din domentul inteligenţei artificiale şi pentru conexiunea minte-corp;

Faţă de Dileep George, pentru ideile lui pătrunzătoare şi pentru discuţiile pline de inspiraţie purtate prin e-mail sau altfel;

Faţă de Martine Rothblatt, pentru devotamentul faţă de toate tehnologiile pe care le discut în această carte şi pentru colaborarea noastră în dezvoltarea tehnologiilor din aceste domenii;

Faţă de echipa KurzweilAI.net, care m-a sprijinit din punctul de vedere al documentării şi logistic în acest proiect, inclusiv faţă de Aaron Kleiner, Amara Angelica, Bob Beal, Casey Beal, Celia Black-Brooks, Cindy Mason, Denise Scutellaro, Joan Walsh, Giulio Prisco, Ken Linde, Laksman Frank, Maria Ellis, Nanda Barker-Hook, Sandi Dube, Sarah Black, Sarah Brangan şi Sarah Reed;

Faţă de echipa devotată de la Viking Penguin, pentru expertiza pe care mi-a pus-o la dispoziţie, inclusiv faţă de Clare Ferraro (preşedinte), Carolyn Coleburn (director de publicitate), Yen Cheong şi Langan Kingsley (agenţi de publicitate), Nancy Sheppard (director de marketing), Bruce Giffords (editor de producţie), Kyle Davis (asistent editorial), Fabiana Van Arsdell (director de producţie), Roland Ottewell (copy editor), Daniel Lagin (designer) şi Julia Thomas (designerul supracopertei);

Faţă de colegii mei de ia Singularity University, pentru ideile, entuziasmul şi energia antreprenorială;

Faţă de colegii care mi-au dat idei generatoare de inspiraţie reflectate în această carte, inclusiv faţă de Barry Ptolemy, Ben Goertzel, David Dalrymple, Dileep George, Felicia Ptolemy, Francis Ganong, George Gilder, Larry

Janowitch, Laura Deming, Lloyd Watts, Martine Rothblatt, Marvin Minsky, Mickey Singer, Peter Diamandis, Raj Reddy, Terry Grossman, Tomaso Poggio şi Vlad Sejnoha;

Faţă de lectorii experţi, inclusiv faţă de Ben Goertzel, David Gamez, Dean Kamen, Dileep George, Douglas Katz, Harry George, Lloyd Watts, Martine Rothblatt, Marvin Minsky, Paul Linsay, Rafael Reif, Raj Reddy, Randal Koene, dr. Stephen Wolfram şi Tomaso Poggio;

Faţă de cititorii mei apropiaţi şi nespecialişti ale căror nume apar mai sus;

Şi, în cele din urmă, faţă de toţi gânditorii creativi din lume care mă inspiră zi de zi.

INTRODUCERE

Creierul – mai mare decât cerul – Căcidacă le-aşezi alături – Unul pe celălalt 11 va cuprinde cu *uşurinţă – cu tine – alături – Creierul e mal adânc decât marea – Căci – uite-le – albastru pe albastru – Unul pe celălalt se înghiţea bureţiicofele – Creierul are. Întocmai greutatea lui Dumnezeu – Căci – cumpăneşte-le – gram cu gram – De s-or deosebi – în caz c-o fi – Ca de sunet, silaba*.

— Emily Dickinson

Fiind cel mai important fenomen din Univers, inteligenţa poate transcende limitele naturale şi poate transforma lumea după propria imagine. În mâinile noastre, inteligenţa ne-a permis să ne depăşim limitările moştenirii biologice şi să ne schimbăm totodată pe noi înşine. Suntem singura specie care face acest lucru.

Povestea inteligenţei omului începe cu un univers capabil să codifice informaţia. Acesta a fost elementul care a facilitat evoluţia. Cum a ajuns Universul să arate astfel este în sine o poveste interesantă. Modelul-standard al fizicii are zeci de constante care trebuie să fie exact cele ce sunt, pentru că, altminteri, atomii n-ar fi fost posibili şi nu ar fi existat nici stele, nici planete, nici creier şi nici cărţi despre creier.

Faptul că legile fizicii sunt atât de precis reglate încât să permită evoluţia informaţiei pare să fie teribil de

improbabil. Şi totuşi, potrivit principiului antropic, nu am vorbi despre ele dacă nu ar fi fost cazul.

Acolo unde unii oameni văd o intervenţie divină, alţii văd un multivers care susţine dezvoltarea de universuri, în timp ce universurile plictisitoare (nepurtătoare de informaţie) mor. Dar, indiferent de felul în care a ajuns

Universul aşa cum este, ne putem începe povestea cu o lume bazată pe informaţie.

Povestea evoluţiei se dezvăluie a fi plină de tot mai multe niveluri de abstracţie. Atomii – în special atomii de carbon, care pot să creeze structuri bogate în informaţie, legându-se pe patru direcţii diferite – au format molecule din ce în ce mai complexe. Ca urmare, fizică a dat naştere chimiei.

Un miliard de ani mai târziu, a apărut o moleculă complexă, numită ADN, care poate să codifice cu precizie şiruri lungi de informaţii şi să genereze organismele descrise de aceste „programe”. Drept rezultat, chimia a dat naştere biologiei.

Cu o rată de dezvoltare tot mai rapidă, organismele şi-au format şi perfecţionat reţele de comunicare şi decizie numite sisteme nervoase, ca să poată coordona părţile din ce în ce mai complexe ale corpurilor lor şi comportamentele care le-au facilitat supravieţuirea. Neuronii care compun sistemele nervoase s-au agregat în creiere capabile de comportamente din ce în ce mai inteligente. În acest fel, biologia a dat naştere neurologiei, creierele fiind în această privinţă vârful de lance al stocării şi manipulării informaţiei. Şi astfel am ajuns de la atomi şi molecule la ADN şi creiere. Următorul pas a fost în mod unic omenesc.

Creierul mamiferelor are o aptitudine distinctă, nemaiîntâlnită la nicio altă clasă de animale. Suntem capabili de gândire *Ierarhică*, de înţelegerea unei structuri compuse din diverse elemente aranjate într-un model, reprezentând acest aranjament cu un simbol şi apoi folosind acel simbol ca element al unei o configuraţii şi mai

complicate. Această capacitate îşi are sediul într-o structură a creierului numită neocortex, care la om a atins un asemenea prag de complexitate, încât suntem capabili să denumim aceste modele *idei*. Printr-un proces recursiv fără sfârşit, suntem în stare să construim idei din ce în ce mai sofisticate. Numim această vastă multitudine de idei legate recursiv între ele *cunoaştere*. Numai *Homo sapiens* are o bază de cunoaştere care evoluează ea însăşi, se dezvoltă exponenţial şi e transmisă de la o generaţie la alta.

Dar creierele noastre au mai dat naştere unui nivel de abstracţie, concretizat în faptul că ne-am folosit inteligenţa, plus încă un factor de potenţare, un apendice opozabil – degetul mare –, pentru manipularea mediului în scopul confecţionării de unelte. Aceste unelte au reprezentat o nouă formă de evoluţie, neurologia dând naştere tehnologiei. Numai datorită uneltelor, baza noastră de cunoaştere a putut să se dezvolte fără limită.

Prima noastră invenţie a fost povestea: limbajul vorbit, care ne-a permis să reprezentăm idei prin enunţuri distincte. Odată cu inventarea ulterioară a limbajului scris, ne-am dezvoltat forme distincte pentru a ne simboliza ideile. Bibliotecile cu limbaj scris au depăşit cu mult capacitatea creierului nostru neasistat de a reţine şi a extinde baza de cunoaştere de idei structurate recursiv.

Există o anumită controversă dacă alte specii, cum şi cimpanzeii, au capacitatea de exprimare a ideilor ierarhice prin limbaj. Cimpanzeii sunt capabili să înveţe un set limitat de simboluri ale limbajului semnelor, pe care le pot folosi pentru comunicarea cu antrenorii umani. Este limpede, pe de altă parte, că există limite clare în ceea ce priveşte complexitatea structurilor de cunoaştere cu care sunt capabili să opereze cimpanzeii. Propoziţiile pe care le pot exprima sunt limitate la anumite secvenţe specifice simple substantiv-verb şi nu sunt capabili de expansiunea nelimitată a complexităţii, caracteristică omului. Ca exemplu distractiv de complexitate a limbajului generat de

om, citeşte una dintre propoziţiile care se întind pe mai multe pagini dintr-o povestire sau nuvelă de Gabriel García Márquez. Povestea întinsă pe şase pagini *Ultima călătorie a fantomei* constă dintr-o singură frază care curge foarte bine atât în limba spaniolă, cât şi în traducerea engleză.1

Ideea fundamentală a celor trei cărţi anterioare ale mele pe tema tehnologiei (*Epoca maşinilor inteligente*, scrisă în anii 1980 şi publicată în 1989; *Epoca maşinilor spirituale*, scrisă la mijlocul anilor 1990 şi publicată în 1999; şi *Singularitatea e aproape*, scrisă la începutul anilor 2000 şi publicată în 2005) este că procesul evolutiv se accelerează în mod inerent (ca urmare a nivelului tot mai ridicat de abstractizare) şi că produsele lui se dezvoltă exponenţial, atât sub aspectul complexităţii, cât şi sub cel al capacităţii. Eu numesc acest fenomen *Legea accelerării veniturilor*, valabilă şi în evoluţia biologică, şi în cea tehnologică. Exemplul cel mai spectaculos de aplicare a Legii accelerării veniturilor îl reprezintă creşterea exponenţială remarcabil de predictibilă a capacităţii şi raportului preţ/performanţă în tehnologia informaţiei. Procesul evolutiv al tehnologiei a condus invariabil la computer, care la rândul lui a făcut posibilă o vastă extindere a bazei noastre de cunoaştere, permiţând realizarea unor conexiuni vaste între un domeniu al cunoaşterii şi altul. Web-ul este intrinsec o expresie elocventă şi adecvată a capacităţii unui sistem ierarhic de a integra o mare cantitate de cunoaştere, în condiţiile păstrării structurii sale inerente. Lumea însăşi este inerent ierarhică. De pildă, copacii au ramuri; ramurile au frunze; frunzele au vinişoare. Clădirile au etaje; etajele au încăperi; încăperile au uşi, ferestre, ziduri şi pardoseli.

De asemenea, am pus la punct unelte care ne permit acum să ne înţelegem propria biologie în termeni de informaţie precişi. Prin retroinginerie, prelucrăm rapid procesele informaţiei aflate la baza biologiei, inclusiv pe cele ale creierului nostru. Avem actualmente codul-obiect al vieţii sub forma genomului uman, o realizare care a

constituit prin ea însăşi un remarcabil exemplu de creştere exponenţială, prin faptul că volumul de date genetice secvenţializate aproape că s-a dublat în lume în fiecare an pe parcursul ultimelor două decenii.2 Astăzi, avem capacitatea de a simula pe computere modul cum dau naştere secvenţele de perechi de baze secvenţelor de aminoacizi ce se adună în proteinele tridimensionale din care este construită întreaga lume biologică. Complexitatea proteinelor pentru care putem simula concentrarea a crescut continuu, pe măsură ce resursele computaţionale au manifestat o creştere exponenţială.3 Putem de asemenea simula modul în care interacţionează proteinele între ele într-un dans tridimensional de forţe atomice extrem de complicat. Înţelegerea tot mai aprofundată a biologiei de către om este una dintre cele mai importante faţete ale descoperirii secretelor inteligente pe care ni le-a transmis evoluţia. Apoi, folosind aceste paradigme inspirate de biologie, urmează crearea unei tehnologii şi mai inteligente.

Actualmente este în curs de derulare un proiect grandios, care antrenează mii şi mii de oameni de ştiinţă şi ingineri. Aceştia se străduiesc să înţeleagă cel mai inteligent exemplu de proces pe care îl avem la dispoziţie: creierul omului. Este în mod justificat cel mai important efort din istoria civilizaţiei om-maşină. În cartea mea *The Singularityls Ne-ar* (*Singularitatea este aproape*), am demonstrat că un corolar al Legii accelerării veniturilor afirmă că s-ar putea să nu existe alte specii inteligente. Ca să rezum demonstraţia: dacă aceste specii ar fi existat, le-am fi observat, dat fiind intervalul de timp relativ redus care se scurge între momentul achiziţiei de către o civilizaţie a tehnologiei brute (să ne gândim că, în 1850, cel mai rapid mijloc de transmitere a informaţiei la nivel naţional, în Statele Unite, era poşta călare – Pony Express) şi momentul deţinerii acelei tehnologii care poate depăşi graniţele propriei planete.4 Din această perspectivă, supunerea creierului uman la mijloacele retroingineriei

poate fi considerată cel mai proiect din Univers.

Scopul proiectului este înţelegerea precisă a modului cum funcţionează creierul uman, urmată de utilizarea metodelor scoase la lumină pentru a ne înţelege mai bine pe noi înşine, pentru repararea creierului la nevoie şi – cel mai relevant în privinţa subiectului acestei cărţi – pentru crearea unor maşini şi mai inteligente. Nu uita, cititorule, că ingineria este capabilă tocmai de amplificarea unui fenomen natural. Ca exemplu, să ne gândim la Principiul lui Bernoulli, care afirmă că există o presiune ceva mai redusă pe o suprafaţă curbă în mişcare, decât pe una dreaptă. Aspectele matematice ale modului în care Principiul lui Bernoulli descrie portanţa aripii nu sunt încă definitiv stabilite în rândurile oamenilor de ştiinţă, dar ingineria a preluat această idee delicată, şi-a concentrat forţele şi a creat întreaga lume a aviaţiei.

În această carte, prezint o teză pe care o numesc Teoria minţii bazate pe recunoaşterea modelelor – *Pattern recognition theory of mind* (*PRIM*) – care, după cum demonstrez, descrie algoritmul de bază al neocortexului (regiunea creierului răspunzătoare de percepţie, memorie şi gândire critică), în capitolele care yrmează, prezint modul în care cercetările recente din domeniul neuroştiinţei, dar şi experimentele noastre mintale, conduc la concluzia inevitabilă că această metodă e folosită total la nivelul neocortexului. Implicaţiile Teoriei minţii bazate pe recunoaşterea modelelor, combinată cu Legea accelerării veniturilor, spun că vom fi capabili să valorificăm prin inginerie aceste principii pentru extinderea extraordinară a puterii propriei noastre inteligenţe.

Iar acest proces este deja în curs de desfăşurare. Există o multitudine de funcţii şi activităţi – care anterior reprezentau apanajul exclusiv al inteligenţei umane – care acum pot fi realizate de computere, de obicei cu mai multă precizie şi pe o scară mult mai mare. De fiecare dată când trimiţi un e-mail sau dai un apel pe un telefon celular, algoritme inteligente rulează în mod optim informaţia. Faci

o electrocardiogramă, iar aceasta îţi este furnizată împreună cu un diagnostic computerizat care rivalizează cu diagnosticele medicilor. Acelaşi lucru este valabil şi în ceea ce priveşte imaginile elementelor figurate ale sângelui. Algoritmi inteligenţi detectează automat fraudele cu carduri bancare, pilotează şi aduc la sol avioane, ghidează sisteme de arme inteligente, ajută la conceperea unor produse prin proiectare asistată de computer şi urmăresc nivelurile de inventar la momentul oportun, asamblează produse în fabrici robotizate şi joacă diverse jocuri, cum ar fi şahul şi chiar subtilul Go, la nivel de maeştri.

Milioane de oameni au fost martorii momentului când un computer IBM pe nume Watson a concurat la *Jeopardyi*, un joc bazat pe limbajul natural, şi a obţinut un scor mai mare decât cei mai buni doi jucători umani din lume la un loc. Ar trebui subliniat nu numai faptul că Watson a citit şi a „înţeles" limbajul subtil al întrebărilor de la *Jeopardy*! (care include fenomene cum ar fi jocurile de cuvinte şi metaforele), ci şi faptul că şi-achiziţionat singur cunoştinţele necesare pentru a da răspunsurile, asimilând sute de milioane de pagini de documente în limbaj natural, inclusiv Wikipedia şi alte enciclopedii. A trebuit să stăpânească practic toate domeniile care ilustrează strădaniile intelectuale ale omului, inclusiv istoria, ştiinţele, literatura, artele, cultura şi multe altele. Compania IBM lucrează actualmente, împreună cu Nuance Speech Technologies (anterior Kurzweil Computer Products, prima mea companie), la o nouă versiune a lui Watson, care va citi literatură medicală (în esenţă, toate jurnalele şi blogurile medicale importante) – pentru a deveni maestru al diagnosticului şi consultant medical – folosind limbajul clinic Nuance: înţelegerea tehnologiilor. Unii observatori au argumentat că, de fapt, Watson nu „înţelege" cu adevărat întrebările de la *Jeopardy*! sau enciclopediile pe care le-a citit şi că realizează doar o „analiză statistică". Un aspect-cheie pe care îl voi descrie

aici priveşte faptul că tehnicile matematice dezvoltate în domeniul inteligenţei artificiale (cum ar fi cele folosite de Watson şi Siri, asistentul iPhone) sunt foarte asemănătoare din punct de vedere matematic cu metodele pe care le-a dezvoltat biologia sub forma neocortexului. Dacă înţelegerea limbajului şi a altor fenomene prin analiză statistică nu explică adevărata înţelegere, atunci nici oamenii nu sunt înzestraţi cu putere de înţelegere.

Capacitatea lui Watson de a stăpâni în mod inteligent cunoaşterea din documente în limbaj natural se va regăsi pe un motor de căutare pe care îl vei avea şi tu la dispoziţie – şi asta chiar în curând. Deja oamenii vorbesc cu telefoanele în limbaj natural (prin intermediul lui Siri, la care a contribuit şi Nuance). Aceşti asistenţi în limbaj natural vor deveni rapid mai inteligenţi, pe măsură ce vor folosi mai multe metode gen Watson. Watson însuşi va continua să se perfecţioneze.

Automobilele Google fără şofer au parcurs deja 320.000 de kilometri în oraşe mari şi mici din California (fără îndoială, cifra va fi cu mult mai mare când această carte va poposi pe rafturile reale şi virtuale). Există multe alte exemple de inteligenţă artificială în lumea de astăzi şi cu mult mai multe se întrezăresc la orizont.

Altă ilustrare a Legii accelerării veniturilor se referă la rezoluţia spaţială a scanării creierului. Volumul de date pe care le culegem despre creier se dublează de la an la an. În plus, demonstrăm că putem transforma aceste date în modele de lucru şi în simulări ale regiunilor creierului. Am reuşit să reproducem prin retroinginerie funcţii-cheie ale cortexului auditiv, unde procesăm informaţiile despre sunete; ale cortexului vizual, unde procesăm informaţii de la analizatorul vizual; şi ale cerebelului, cu care contribuim parţial la formarea abilităţilor (cum ar fi prinderea unei mingi).

În avangarda proiectului de înţelegere, modelare şi simulare a creierului uman stă intenţia de a reproduce prin retroinginerie neocortexul cerebral, în care realizăm

gândirea ierarhică recursivă. Cortexul cerebral reprezintă 80 la sută din creierul omului şi este format din structuri extrem de repetitive, care le permit oamenilor să creeze ansambluri de idei cu complexităţi arbitrare.

În cadrul Teoriei minţii pe baza recunoaşterii modelelor, descriu un model de achiziţionare de către creierul uman a acestei capacităţi fundamentale folosind o structură foarte inteligentă, proiectată de evoluţia biologică. În acest mecanism cortical există detalii pe care încă nu le înţelegem pe deplin, dar cunoaştem suficient de bine funcţiile pe care trebuie să le realizeze acesta pentru a putea să creăm algoritmi care să corespundă exact aceluiaşi scop. Începând să înţelegem neocortexul, ne aflăm acum în poziţia de a-i amplifica foarte mult puterea, la fel cum lumea aviaţiei a amplificat foarte mult puterea Principiului lui Bernoulli. Principiul de operare al neocortexului este în mod demonstrabil cea mai importantă idee din lume, deoarece e capabil să reprezinte integralitatea cunoaşterii şi aptitudinilor şi să creeze noi cunoştinţe. La urma urmelor, neocortexul a fost răspunzător de toate romanele, cântecele, picturile, descoperirile ştiinţifice, ca şi de alte, felurite, produse ale gândirii omeneşti.

În domeniul neuroştiinţei, este foarte necesară o lege care să facă legătura între observaţiile extrem de disparate şi extinse comunicate zilnic. O teorie unificată reprezintă cerinţa crucială în toate domeniile ştiinţei. În Capitolul 1, voi arăta cum doi visători au unificat biologia şi fizica, domenii care anterior păruseră disperant de dezordonate şi străine unul de altul, iar pe urmă mă voi ocupa de modul în care o asemenea teorie poate fi aplicată peisajului creierului.

Astăzi întâlnim adeseori grandioase manifestări ale complexităţii creierului uman. Google dă în jur de 30 de milioane de link-uri la o căutare care solicită citate pe această temă. (Este imposibil, însă, să se traducă asta în numărul real de citate pe care le afişează, deoarece unele

site-uri web conţin citate multiple, pe când altele nu conţin niciunul.) James D. Watson însuşi a scris în 1992: „Creierul este ultima şi cea mai măreaţă frontieră a biologiei, cel mai complex lucru care a fost descoperit până acum în Universul nostru”.

El continuă şi explică de ce crede că „[creierul] conţine sute de miliarde de celule interconectate prin intermediul a mii de miliarde de conexiuni. Creierul descumpăneşte mintea”.5 Sunt de acord cu sentimentul lui Watson referitor la faptul că creierul este cea mai măreaţă frontieră a biologiei, însă faptul că acesta conţine miliarde de celule şi mii de miliarde de conexiuni nu înseamnă că metoda lui principală este foarte complexă, dacă putem identifica modele uşor de înţeles (şi recreabile) în aceste celule şi conexiuni, mai ales în cele masiv redundante.

Să ne gândim la ce înseamnă să fii complex. Am putea să ne întrebăm: O *pădure e complexă*? Răspunsul depinde de perspectiva pe care alegem să o urmăm. Putem observa că în pădure sunt mii de copaci şi că fiecare dintre ei e diferit. Putem să mergem, pe urmă, mai departe şi să observăm că fiecare copac are mii de ramuri şi că fiecare ramură e total diferită de celelalte. Apoi, am putea trece la descrierea ciudăţeniilor unei singure ramuri. Concluzia poate fi aceea că pădurea prezintă o complexitate care depăşeşte şi cea mai prolifică imaginaţie.

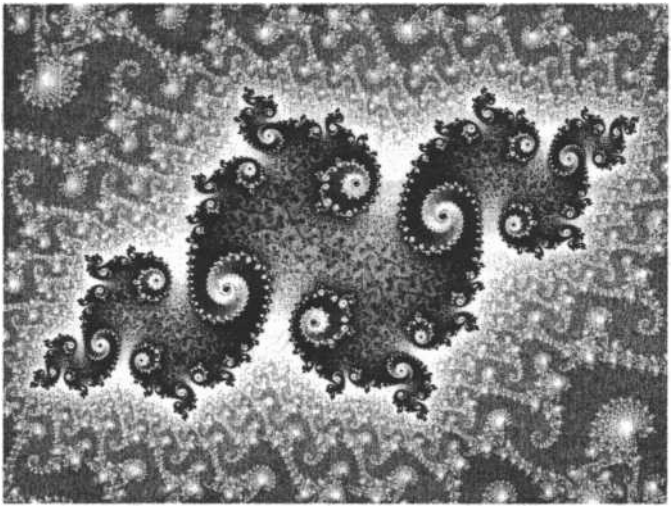
Însă o asemenea abordare ar însemna literalmente să nu vedem pădurea din cauza copacilor. Desigur, există o imensă variaţie fractală între copaci şi între ramuri, dar, pentru a înţelege corect principiile unei păduri, ar fi mai bine dacă am începe cu identificarea modelelor distincte de redundanţă cu variaţie stohastică (respectiv aleatorie) ce pot fi găsite în ea. Ar fi corect să spunem că acest concept de pădure este mai simplu decât conceptul de copac.

La fel stau lucrurile şi în cazul creierului, care prezintă, de asemenea, redundanţe enorme, mai ales la nivelul neocortexului. Aşa cum voi arăta în această carte,

corect ar fi să existe o complexitate mai mare într-un singur neuron, decât în structura de ansamblu a neocortexului.

În mod cert, scopul meu în lucrarea de faţă nu este acela de a adăuga un nou citat la milioanele care există deja şi care atestă cât de complex este creierul; dimpotrivă, doresc să îţi imprim, cititorule, forţa simplităţii sale. Şi voi face acest lucru descriind modul în care un mecanism de bază ingenios, de recunoaştere, reţinere şi predicţie, repetat în neocortex de sute de milioane de ori, explică marea diversitate a gândirii noastre. La fel cum, din diferitele combinaţii ale valorilor codului genetic din ADN-ul nuclear şi mitocondrial, apare o uluitoare diversitate de organisme, tot aşa se formează şi o gamă uluitoare de idei, gânduri şi abilităţi pe baza valorilor modelelor (de conexiuni şi tării sinaptice) care se găsesc m şi între identificatorii noştri neocorticali de modele. Aşa cum spune specialistul în neuroştiinţă de la MIT Sebastián Seung, „identitatea nu stă în genele noastre, ci în conexiunile dintre celulele creierului”.6

Trebuie să facem distincţia între adevărata complexitate de concepţie şi complexitatea aparentă. Să luăm celebrul set Mendelbrot, a cărui imagine este de multă vreme simbolul complexităţii. Pentru înţelegerea caracterului său aparent complicat, este suficient să mărim imaginea (pe care o poţi accesa, cititorule, prin intermediul link-urilor din această notă de final).7 Există o nesfârşită complexitate în complexitate şi toate sunt întotdeauna diferite. Cu toate acestea, concepţia – formula – pentru setul Mendelbrot nu ar putea fi mai simplă de-atât. Are o lungime de şase caractere: $Z = Z_2 + C$, unde Z e un număr „complex” (ceea ce înseamnă o pereche de numere), iar C este o constantă. Nu-i necesar să înţelegi funcţia Mendelbrot pentru a pricepe că e simplă.



Vedere a unei reprezentări a setului Mendelbrot, o formulă simplă care este aplicată recurent. Pe măsura măririi reprezentării pe ecran, imaginile se schimbă în permanenţă în moduri aparent complexe

Această formulă este aplicată recurent la fiecare nivel al unei ierarhii. Lucrul e valabil şi pentru creier. Structura lui repetitivă nu este la fel de simplă precum cea a formulei compuse din şase caractere a setului Mandelbrot, dar nu e nici pe departe atât de complexă pe cât ar sugera-o milioanele de citate referitoare la complexitatea creierului. Acest model neocortical se repetă la nesfârşit la toate nivelurile ierarhiei conceptuale reprezentate de neocortex. Einstein a articulat bine scopurile cărţii de faţă când a spus că „orice nebun inteligent poate să facă lucrurile mai mari şi mai complexe… dar e nevoie… de mult curaj pentru a înainta în direcţia opusă”.

Până acum am vorbit despre creier. Dar cum rămâne cu mintea? De exemplu, cum poate să ajungă la conştiinţă un neocortex care rezolvă probleme? Şi pentru că tot suntem la acest subiect, câte minţi conştiente avem în creierul nostru? Există mărturii care sugerează că e posibil să fie mai mult de una.

O altă întrebare pertinentă referitoare la minte sună

aşa: Ce *este liberul arbitru – şi îl avem noi oare*? Există experimente care par să indice faptul că începem să ne implementăm deciziile încă înainte de a ne da seama că le-am luat deja. Asta să presupună faptul că liberul arbitru e o iluzie?

În sfârşit, care atribute ale creierului nostru sunt răspunzătoare pentru constituirea identităţii proprii? Sunt aceeaşi persoană care eram în urmă cu şase luni? în mod evident, nu mai sunt exact cum eram atunci, dar am aceeaşi identitate?

Vom trece în revistă ce presupune Teoria pentru minte a recunoaşterii modelelor în privinţa acestor întrebări vechi de secole.

Capitolul 1

EXPERIMENTE MINTALE DESPRE LUME

Teoria selecţiei naturale a lui Darwin a apărut foarte târziu în istoria gândirii. O fi fost oare întârziată pentru că se opunea dezvăluirii adevărului, pentru că era o temă cu totul nouă în istoria ştiinţei, pentru că era caracteristică numai fiinţelor vii sau pentru că se ocupa cu scopul şi cauzele finale fără să postuleze un act al creaţiei? Eu cred că nu. Darwin pur şi simplu a descoperit rolul selecţiei, un gen de cauzalitate foarte diferit de mecanismele cauză-efect ale ştiinţei de până în acel moment.

Originile unei fantastice varietăţi de lucruri vii pot fi explicate prin contribuţia acelor caracteristici noi, posibil de provenienţă aleatorie, care au reuşit să supravieţuiască. În ştiinţele fizicii şi biologiei exista prea puţin sau nu exista nimic care să prevestească selecţia drept cauză principală.

— B.F. Skinner în ultimă instanţă, nimic nu este sacru, cu excepţia integrităţii propriei minţi.

— Ralph Waldo Emerson

O *metaforă din geologie*

La începutul secolului al nouăsprezecelea, geologii au meditat la o întrebare fundamentală. Peşterile şi canioanele de felul Marelui Canion din Statele Unite şi al Defileului Vikos din Grecia (recunoscut ca fiind cel mai

adânc canion din lume) există în toate colţurile lumii. Cum au ajuns acolo aceste formaţiuni maiestuoase?

Invariabil, a existat un curs de apă care a părut să profite de oportunitatea de a curge prin aceste structuri naturale, dar, anterior mijlocului de secol nouăsprezece, păruse absurd ca aceste cursuri de apă blânde să fi fost creatoarele unor văi şi promontorii atât de mari. Însă geologul britanic Charles Lyell (1797 – 1875) a avansat ipoteza că într-adevăr deplasarea apei a fost aceea care a sculptat aceste modificări geologice majore, de-a lungul unor perioade îndelungate, în esenţă grăunte cu grăunţe. Ideea lui Lyell a fost iniţial întâmpinată cu ridiculizări, dar, în decurs de două decenii, teza a căpătat susţinere generală.

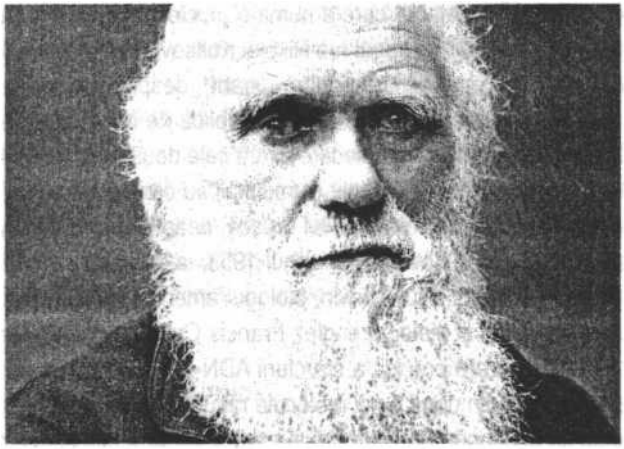
Un om care urmărea cu atenţie răspunsul comunităţii ştiinţifice la teza radicală a lui Lyell a fost naturalistul englez Charles Darwin (1809 – 1882). Să ne gândim la situaţia biologiei în jurul anului 1850. Domeniul era extraordinar de stufos, se confrunta cu nenumărate specii de animale şi plante, oricare dintre ele prezentând o mare complexitate. Majoritatea oamenilor de ştiinţă s-au opus oricărei încercări de a veni cu o teorie unificatoare a uluitoarei varietăţi a Naturii. Această diversitate slujea drept testament pentru glorificarea creaţiei lui Dumnezeu, ca să nu mai vorbim de inteligenţa oamenilor de ştiinţă capabili să o stăpânească.

Darwin a abordat problema conceperii unei teorii generale a speciilor făcând analogie cu teza lui Lyell pentru explicarea modificărilor graduale ale caracterişticilor speciilor pe parcursul mai multor generaţii. El a combinat această idee cu propriile sale experimente mintale şi cu observaţiile de pe parcursul celebrei lui lucrări *Călătoria corăbiei Beagle*. Darwin a demonstrat că, de-a lungul fiecărei generaţii, indivizii care pot supravieţui cel mai bine în nişa lor ecologică ar fi şi cei care ar crea următoarea generaţie.

În data de 22 noiembrie 1859, cartea lui Darwin

Originea speciilor a fost pusă în vânzare. În ea, autorul îşi exprima clar datoria faţă de Lyell:

Sunt perfect conştient că această doctrină a selecţiei naturale, exemplificată în anterioarele posturi imaginare, este expusă aceloraşi obiecţii care au fost emise la început şi împotriva nobilelor perspective ale lui Sir Charles Lyell cu privire la „modificările moderne ale Pământului, ilustrate de geologie"; dar acum rareori mai auzim că acţiunea valurilor de pe coaste mai este considerată o cauză superficială şi nesemnificativă atunci când e aplicată excavării unor văi gigantice sau formării celor mai lungi linii de promontorii din interiorul uscatului. Selecţia naturală nu poate să intervină decât prin acumularea şi păstrarea unor modificări infinit de mărunte, fiecare dintre ele profitabilă pentru fiinţa actuală; şi, la fel cum geologia modernă aproape că a alungat concepţii cum ar fi excavarea unei mari văi de un singur val de proporţii diluviene, tot aşa selecţia naturală, dacă este un principiu adevărat, va elimina credinţa în crearea permanentă de noi fiinţe organice sau în orice mari şi bruşte modificări ale structurii lor.1



Charles Darwin, autorul cărţii *Originea speciilor*, care a consacrat ideea de evoluţie biologică întotdeauna există mai multe motive pentru care se manifestă opoziţie faţă de marile idei noi şi nu este greu să identificăm asemenea motive şi în cazul lui Darwin. Afirmaţia că ne-am

tras nu de la Dumnezeu, ci din maimuţe, iar înaintea acestora din viermi, nu Le-a căzut bine multor comentatori. Aluzia că, bunăoară, câinele nostru de companie ne este văr, la fel ca şi coropişniţa, ca să nu mai vorbim de planta pe care am călcat (văr de gradul un milion sau un miliard, poate, dar totuşi rudă) li s-a părut multora o blasfemie.

Însă ideea a prins rapid, pentru că aducea coerenţă în ceea ce anterior reprezentase un corp de observaţii aparent fără legătură între ele. Înainte de anul 1872, la publicarea celei de-a şasea ediţii a cărţii *Originea speciilor*, Darwin a adăugat într-un pasaj: „Ca o consemnare a unei stări anterioare de lucruri, am păstrat în paragrafele precedente… mai multe fraze care sugerează că naturaliştii cred într-o creaţie separată a fiecărei specii; şi am fost foarte cenzurat pentru că m-am exprimat astfel. Dar, fără îndoială, aceasta era credinţa generală atunci când a apărut prima ediţie a prezentei lucrări. Acum lucrurile s-au schimbat complet şi aproape orice naturalist recunoaşte marele principiu al evoluţiei”.2

Pe parcursul următorului secol, ideea unificatoare a lui Darwin s-a adâncit. În anul 1869, la doar un deceniu de la publicarea iniţială a *Originii speciilor*, fizicianul elveţian Friedrich Miescher (1844 – 1895) a descoperit în nucleul celulei o substanţă pe care a numit-o „nucleină” şi care s-a dovedit a fi ADN-ul.3 În anul 1927, biologul rus Nikolai Koltsov (1872 – 1940) a descris cecă ce el numea o „moleculă ereditară gigant”, despre care a spus că e formată din „două şiruri în oglindă, care se replică de o manieră semiconservatoare, folosind drept model fiecare dintre cele două şiruri”. Şi descoperirea lui a fost condamnată de mulţi. Comuniştii au considerat că era vorba de propagandă nazistă, iar moartea lui bruscă, neaşteptată, a fost atribuită poliţiei secrete a Uniunii Sovietice.4 În anul 1953, la aproape un secol de la publicarea prolificei cărţi a lui Darwin, biologul american James D. Watson (născut în anul 1928) şi biologul englez

Francis Crick (1916 – 2004) au realizat prima caracterizare precisă a structurii ADN-ului, descriindu-l drept un dublu helix compus din două lungi molecule răsucite.5 Merită să scot în evidenţă faptul că descoperirea celor doi s-a bazat pe ceea ce actualmente e cunoscut sub numele de „Fotografia 51”, realizată de colegalor, Rosalind Franklin, care a folosit cristalografia în raze X. A fost prima reprezentare care a dezvăluit dublul helix. Date fiind indiciile obţinute din fotografia lui Franklin, au fost emise sugestii potrivit cărora Rosalind Franklin ar fi trebuit să primească, alături de Watson şi Crick, Premiul Nobel.

Rosalind Franklin a realizat fotografia esenţială a ADN-ului (folosind cristalografia în raze X), care le-a permis lui Watson şi Crick să descrie cu acurateţe, pentru prima oară, structura ADN-ului

Odată cu descrierea moleculei care putea să codifice programul biologic, a fost instaurată şi o teorie unificatoare a biologiei. Aceasta a asigurat un fundament simplu şi elegant pentru toate formele de viaţă. Depinzând numai de valorile perechilor de bază care alcătuiesc şirurile ADN-ului din nucleu (şi, într-o mai mică măsură, din mitocondrii), organismul devine la maturitate un fir de

iarbă sau o fiinţă omenească. Intuirea acestui fapt nu a eliminat încântătoarea diversitate a Naturii, dar acum înţelegem că această extraordinară diversitate izvorăşte din marea varietate de structuri care pot fi codificate în această moleculă universală.

Călare pe o rază de lumină

La începutul secolului douăzeci, lumea fizicii se afla într-o fundătură, în urma unei noi serii de experimente. În anul 1879, în familia unui inginer german şi a soţiei sale casnice s-a născut un băiat. A început să vorbească abia la vârsta de trei ani şi s-a spus despre el că, la nouă ani, a avut probleme la şcoală. La şaisprezece ani, visa cu ochii deschişi să călărească pe o rază de lumină.

Acest băiat era conştient de experimentul din anul 1803 al matematicianului englez Thomas Young (1773 – 1829), care a stabilit că lumina e formată din unde. Concluzia, la momentul acela, a fost că undele de lumină trebuie să se deplaseze printr-un fel de mediu; la urma urmelor, valurile oceanului se propagă prin apă, iar undele sonore se propagă prin aer şi alte materiale. Oamenii de ştiinţă au denumit mediul prin care se propagă lumina *eter*. Băiatul era la curent şi cu experimentul din anul 1887 al savanţilor americani Albert Michelson (1852 – 1931) şi Edward Morley (1838 – 1923), care încercaseră să confirme existenţa eterului. Acel experiment se bazase pe analogia deplasării unei bărci cu vâsle în amonte şi în aval pe un râu. Dacă vâsleşti cu viteză constantă, atunci viteza ta măsurată de pe mal va fi mai mare când vâsleşti în aval decât când vâsleşti în amonte. Michelson şi Morley au presupus că lumina se deplasează prin eter cu o viteză constantă (respectiv cu viteza luminii). Ei au raţionat că viteza luminii Soarelui atunci când Pământul se deplasează pe orbita lui spre Soare (viteză măsurată din punctul nostru de vedere, de pe Pământ) comparativ cu viteza ei aparentă când Pământul se depărtează de Soare trebuie să fie diferită (cu dublul vitezei Pământului). Dovedirea acestui lucru ar fi confirmat existenţa eterului. Însă ceea

ce au descoperit cei doi americani a fost faptul că nu există nicio diferenţă între vitezele cu care lumina solară ajunge pe Pământ, indiferent unde se află acesta pe orbita lui. Descoperirea infirma ideea de eter, dar ce se întâmpla în realitate? Acest lucru a rămas un mister timp de aproape două decenii.

În timp ce îşi imagina că se plimbă de-a lungul unei raze, adolescentul german s-a gândit că trebuia să vadă razele de lumină nemişcate, la fel cum un tren ar părea că nu se mişcă deloc dacă te-ai deplasa alături de el cu aceeaşi viteză. Însă şi-a dat seama că acest lucru era imposibil, deoarece se presupunea că viteza luminii e constantă indiferent de mişcarea observatorului. Prin urmare, şi-a imaginat că se deplasează de-a lungul unei raze de lumină, dar la o viteză întrucâtva mai mică. Şi dacă se deplasa cu 90 la şută din viteza luminii? Dacă razele de lumină sunt asemenea trenurilor, a raţionat el, atunci trebuie să vadă raza de lumină deplasându-se cu 10 la sută din viteza luminii. Şi, într-adevăr, asta ar fi trebuit să vadă observatorii de pe Pământ. Însă ştim că viteza luminii este o constantă, aşa cum arătase experimentul Michelson-Morley. Prin urmare, tânărul ar fi văzut în mod necesar raza de lumină deplasându-se înaintea lui cu întreaga viteză a luminii. Acest fapt părea o contradicţie – deci cum putea fi posibil?

Răspunsul i-a devenit evident germanului, al cărui nume era, întâmplător, Albert Einstein (1879 – 1955), înainte să împlinească douăzeci şi şase de ani. În mod evident – pentru tânărul maestru Einstein – *însuşi timpul ar fi trebuit să încetinească pentru el*. Einstein îşi explică raţionamentul într-o lucrare publicată în anul 1905.7 Dacă observatorii de pe Pământ ar fi putut să se uite la ceasul tânărului, l-ar fi văzut ticăind de zece ori mai încet. Şi pe bună dreptate, la întoarcerea pe Pământ, ceasul lui ar fi arătat că s-au scurs numai 10 procente din timp (ignorând, pentru moment, accelerarea şi deceierarea). Din perspectiva lui, însă, ceasul ar funcţiona normal şi raza de

lumină de lângă el s-ar deplasa cu viteza luminii. Încetinirea de zece ori a vitezei timpului însuşi (relativ la ceasurile de pe Pământ) explică pe deplin aparentele discrepanţe de perspectivă. La extrem, încetinirea trecerii timpului ar ajunge la zero în momentul în care viteza deplasării ar atinge viteza luminii; prin urmare, era imposibil să se deplaseze odată cu raza de lumină. Cu toate că deplasarea cu viteza luminii era imposibilă, s-a dovedit a nu fi teoretic imposibilă deplasarea cu viteze *superioare* celei a razei de lumină, în acest caz, timpul ar curge invers.

Acest deznodământ li s-a părut absurd multora dintre criticii iniţiali. Cum putea timpul să îşi încetinească curgerea doar în funcţie de viteza deplasării cuiva? Şi, într-adevăr, timp de optsprezece ani de la momentul experimentului Michelson-Morley, alţi gânditori nu reuşiseră să găsească o concluzie care i se părea atât de evidentă Maestrului Einstein. Acei mulţi alţii care cumpăniseră această problemă în ultima parte a secolului al nouăsprezecelea practic „căzuseră de pe cal” în privinţa urmăririi implicaţiilor unui principiu, ţinându-se în schimb de ideile lor preconcepute privitoare la mecanismul realităţii. (Probabil că ar fi trebuit să modific metafora în „căzuseră de pe raza de lumină”.)

Al doilea experiment mintal al lui Einstein a fost să se imagineze pe el şi pe fratele său zburând prin spaţiu. Cei doi se află la 297.000 de kilometri distanţă unul de altul. Einstein vrea să se deplaseze mai repede, dar doreşte şi să păstreze distanţa dintre ei constantă. Aşadar îi face fratelui său un semn cu lanterna de fiecare dată când vrea să accelereze. Deoarece ştie că va fi nevoie de o. secundă pentru ca semnalul să ajungă la fratele lui, aşteaptă timp de o secundă după emiterea semnalului înainte să efectueze propria accelerare. De fiecare dată când recepţionează semnalul, fratele lui accelerează imediat. În acest fel, cei doi fraţi accelerează exact în acelaşi timp şi, prin urmare, rămân la aceeaşi distanţă unul de celălalt.

Dar acum să ne gândim la ce am vedea dacă ne-am afla pe Pământ. Dacă fraţii s-ar îndepărta de noi (Albert fiind în faţă), ni s-ar părea că ar dura mai puţin de o secundă pentru ca lumina să ajungă la fratele lui, pentru că el se îndreaptă către sursa de lumină. De asemenea, am vedea ceasul fratelui lui Albert încetinind (pe măsură ce viteza lui creşte şi el se află mai aproape de noi). Din ambele motive, i-am vedea pe amândoi fraţii apropiindu-se din ce în ce mai mult şi în cele din urmă ciocnindu-se. Însă, din perspectiva lor proprie, cei doi fraţi ar rămâne la o distanţă constantă de 297.000 de kilometri unul de celălalt.

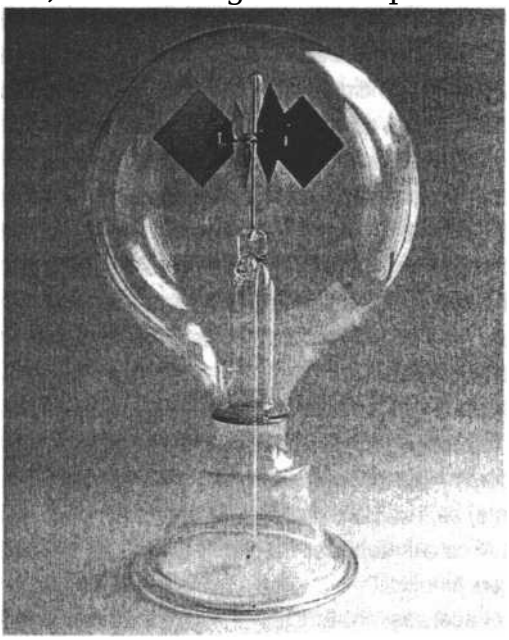
Cum e posibil aşa ceva? Răspunsul – *evident* – este că distanţele se contractă în paralel cu deplasarea (dar nu şi perpendicular pe ea). Prin urmare, cei doi fraţi Einstein ar deveni mai scurţi (presupunând că ar zbura cu capul înainte) pe măsură ce s-ar deplasa tot mai rapid. Această concluzie bizară probabil că a îndepărtat mai mulţi fani iniţiali ai lui Einstein decât diferenţa trecerii timpului.

În cursul aceluiaşi an, Einstein s-a gândit ia relaţia dintre materie şi energie. Era un alt experiment mintal. Fizicianul scoţian James Clerk Maxwell arătase în anii 1850 că particulele de lumină, numite fotoni, nu aveau masă, dar cu toate acestea aveau impuls. Când eram mic, am avut un dispozitiv numit radiometru Crookes8, care consta dintr-o sferă de sticlă închisă ermetic, în care era un vid parţial şi un set de patru paie ce se învârteau pe un ax. Palele erau albe pe o parte şi negre pe cealaltă. Faţa albă a fiecărei pale reflecta lumina, iar faţa neagră o absorbea. (De-asta e mai răcoritor să porţi un tricou alb şi nu unul negru într-o zi caniculară.) Când asupra dispozitivului se aplică un fascicul de lumină, palele se roteau, feţele negre depărtându-se de lumină. Aceasta este o demonstraţie directă că fotonii posedă un impuls suficient de mare pentru a face paiele radiometrului să se mişte.9

Problema cu care s-a luptat Einstein a fost că

impulsul reprezintă o funcţie de masă: impulsul este egal cu masa de înmulţit cu viteza. Astfel, o locomotivă care se deplasează cu 48 de kilometri pe oră are un impuls mult mai mare decât, să spunem, o insectă care se deplasează cu aceeaşi viteză. Şi atunci cum poate o particulă cu masa zero să aibă un impuls pozitiv?

Experimentul mintal al lui Einstein a luat în calcul o cutie care pluteşte în spaţiu. În interiorul cutiei este lansat un foton, de la stânga la dreapta.



Un radiometru Crookes – girueta cu patru pale se roteşte când asupra ei se proiectează lumină

Impulsul total al sistemului trebuie să se conserve, astfel încât cutia ar fi de aşteptat să reculeze spre stânga la emiterea fotonului. După un anumit interval de timp, fotonul se ciocneşte de faţa din dreapta a cutiei, transferându-i acesteia impulsul înapoi. Impulsul total al sistemului se conservă din nou, astfel încât cutia nu se mai mişcă.

Toate bune până aici. Dar să ne gândim din perspectiva punctului de observaţie al domnului Einstein, care priveşte cutia din afară. Nu vede nicio influenţă

exterioară asupra cutiei. Nicio particulă – cu sau fără masă – nu loveşte cutia şi nimic nu iese din ea. Cu toate acestea, domnul Einstein, potrivit scenariului de mai înainte, vede cutia mişcându-se trecător la stânga şi apoi oprindu-se. Potrivit analizei noastre, toţi fotonii ar trebui să deplaseze cutia permanent spre stânga. Cum n-a existat niciun efect exterior asupra cutiei şi nici vreunul generat din cutie, centrul de masă ai acesteia trebuie să rămână în acelaşi loc, deoarece fotonul dinăuntrul cutiei, care se deplasează de la stânga la dreapta, nu poate schimba centrul de greutate, deoarece nu are masă.

Sau poate are? Concluzia lui Einstein a fost că, de vreme ce fotonul în mod evident are energie şi impuls, trebuie să aibă şi un echivalent al masei. Energia fotonului aflat în mişcare este echivalentă întru totul cu a unei mase în mişcare. Putem calcula echivalenţa admiţând că centrul de greutate al sistemului trebuie să rămână staţionar în timpul mişcării fotonului. Realizând suportul matematic, Einstein a arătat că masa şi energia sunt echivalente şi sunt legate de o constantă simplă. Pe de altă parte, exista o chichiţă: constanta poate că era simplă, dar s-a dovedit a fi enormă; era viteza luminii ia pătrat (în jur de 1, 7 x 1017 metri pătraţi pe secundă la pătrat – respectiv 17 urmat de 16 zerouri). De aici obţinem celebra formulă a lui Einstein $E = mc^2$.10 Prin urmare, o uncie (28 de grame) de masă este echivalentă cu 600.000 de tone de trinitrotoluen. Scrisoarea lui Einstein din data de 2 august 1939 adresată preşedintelui Roosevelt, prin care îl informa despre potenţialul unei bombe atomice bazate pe această formulă, a inaugurat epoca atomică.11

Ai crede că acest lucru trebuia să fie evident de mai mult timp, dat fiind faptul că experimentatorii observaseră că masa substanţelor radioactive se reducea în timp, ca rezultat al radiaţiei. Dar s-a presupus că substanţele radioactive conţineau un combustibil special de înaltă energie care ardea. Această ipoteză nu este complet greşită; doar că acel combustibil „ars” era, pur şi simplu,

masa.

Există mai multe motive pentru care am început această carte cu experimentele mintale ale lui Darwin şi Einstein. În primul rând, ele demonstrează extraordinara forţă a creierului uman. Fără niciun echipament în afara unui creion şi a unei hârtii, pentru desenarea prin linii a personajelor din aceste experimente mintale simple şi pentru scrierea ecuaţiilor cât se poate de simple care rezultă din ele, Einstein a reuşit să răstoarne o înţelegere a lumii fizice care data din urmă cu două secole, să influenţeze profund cursul istoriei (inclusiv al Doilea Război Mondial) şi să inaugureze era nucleară.

Este adevărat, Einstein s-a bazat pe câteva descoperiri experimentale ale secolului al nouăsprezecelea, cu toate că nici în aceste experimente nu s-a folosit vreun echipament special. Este de asemenea adevărat că, în validările experimentale ulterioare ale teoriilor lui Einstein, s-au folosit tehnologii avansate şi că, dacă acestea nu ar fi fost concepute, nu am beneficia de confirmarea pe care o avem astăzi la dispoziţie că ideile lui Einstein sunt autentice şi importante. Pe de altă parte însă, asemenea factori nu umbresc cu nimic faptul că aceste faimoase experimente mintale dezvăluie forţa gândirii umane la nivelul ei cel mai înalt.

Einstein este larg recunoscut drept cel mai însemnat om de ştiinţă al secolului al douăzecilea (şi Darwin ar fi un bun pretendent la această onoare pentru secolul nouăsprezece), deşi suportul matematic care se află la baza teoriilor lui nu este, în ultimă instanţă, deosebit de complicat. Iar experimentul mintal propriu-zis a fost cât se poate de simplu. Şi atunci am putea să ne întrebăm din ce punct de vedere ar trebui să-l considerăm pe Einstein deosebit de deştept? Vom discuta mai târziu despre ce anume a făcut el mai exact în creierul lui până să vină cu aceste teorii şi unde anume sălăşluieşte această calitate.

Pe de altă parte, o asemenea istorie demonstrează şi limitele gândirii omeneşti. Einstein a putut călări pe raza

de lumină fără să cadă de pe ea (deşi a concluzionat că era imposibil de fapt să călărească o rază de lumină), însă câte mii de observatori şi gânditori n-au fost absolut lipsiţi de capacitatea de a gândi asemenea exerciţii remarcabil de necomplicate? Un neajuns obişnuit îl reprezintă dificultatea cu care se confruntă majoritatea oamenilor în privinţa eliminării şi depăşirii ideilor şi a perspectivelor semenilor lor. Şi mai există şi alte neajunsuri, pe care le vom discuta în detaliu după ce examinăm felul cum funcţionează neocortexul.

Un model unificat al neocortexului

Cel mai important motiv pentru care îţi împărtăşesc, cititorule, poate cele mai faimoase experimente mintale din istorie este acela că ele constituie o introducere în utilizarea aceleiaşi abordări la analizarea creierului. După cum vei vedea, putem ajunge remarcabil de departe în desluşirea modului cum funcţionează inteligenţa umană prin intermediul câtorva simple experimente mintale proprii. Dacă avem în vedere materia implicată, experimentele mintale ar trebui să constituie o abordare cât se poate de potrivită.

Dacă gândurile leneşe ale unui tânăr şi folosirea unui echipament care nu a depăşit un creion şi o hârtie au fost suficiente pentru a revoluţiona înţelegerea fizicii, atunci ar trebui să fim în măsură să realizăm un progres remarcabil recurgând la un fenomen cu care suntem mult mai familiarizaţi. La urma urmelor, cu toţii trăim experienţa gândirii în fjecare clipă a vieţii noastre – în starea de veghe şi, de asemenea, în lumea viselor.

După ce vom construi un model al felului cum se petrec lucrurile prin acest proces de autoreflecţie, vom examina măsura în care îl putem confirma prin intermediul celor mai noi observaţii realizate pe creier real şi prin recrearea, la nivel de ultimă oră, a acestor procese cu ajutorul maşinilor.

Capitolul 2

EXPERIMENTELE MINTALE DE GÂNDIRE

Foarte rar se întâmplă să gândesc în cuvinte. Am câte un gând şi pot încerca să-l exprim prin cuvinte după aceea.

— Albert Einste în

Creierul este o masă de un kilogram şi jumătate pe care o poţi ţine în mână şi care poate să conceapă un univers cu diametrul de o sută de miliarde de ani-lumină.

— Marian Diamond

Ce mi se pare uluitor este faptul că un simplu obiect de un kilogram şi jumătate, compus din aceiaşi atomi care alcătuiesc toate celelalte lucruri de sub Soare, este capabil să conducă absolut tot ce a făcut omul: să zboare pe Lună şi să înscrie şaptezeci de eseuri, să scrie *Hamlet* şi să construiască Taj Mahal-ul – ba chiar şi să dezlege secretele creierului însuşi.

— Joel Havemann

Am început să gândesc asupra gândirii în jurul anului 1960, acelaşi an în care am descoperit computerul. Astăzi ţi-ar fi foarte greu să găseşti un copil de doisprezece ani care să nu folosească un computer, dar pe vremea aceea erau doar câţiva în această situaţie în oraşul meu natal, New York. Desigur, aceste dispozitive timpurii nu puteau fi ţinute în mână şi primul la care am avut acces ocupa o încăpere mare. La începutul anilor 1960, am făcut nişte tentative de programare pe un computer IBM 1620 pentru analize de multiplicitate (un test statistic), cu date care fuseseră adunate prin studierea unui program pentru educaţia la vârste mici – un precursor al programului *Head Start*. Prin urmare, efortul presupunea un dramatism considerabil, soarta acestei iniţiative educative la nivel naţional depinzând de lucrarea noastră. Algoritmii şi datele analizate erau suficient de complexe încât să nu putem anticipa răspunsurile pe care urma să le dea computerul. Bineînţeles, răspunsurile erau determinate de date, dar nu erau predictibile.

S-a dovedit că distincţia dintre *determinate* şi *predictibile* este importantă, lucru la care voi reveni.

Îmi amintesc cât de incitant a fost momentul când

luminiţele de pe panoul frontal s-au estompat, exact înainte ca algoritmul să-şi încheie deliberările, ca şi cum computerul ar fi fost adâncit în gânduri. Când veneau diferite persoane, nerăbdătoare să primească următorul set de rezultate, le arătam cu degetul luminiţele care licăreau slab şi le spuneam: „Gândeşte”. Era şi nu era o glumă – chiar părea că ar fi cântărit răspunsurile –, iar personalul a început să-i atribuie maşinii o personalitate. Era o antropomorfizare, poate, dar ea m-a făcut să încep să mă gândesc serios la relaţia dintre gândire şi computare.

Pentru a evalua măsura în care propriul meu creier se asemăna cu programele de computer familiare mie pe atunci, am început să meditez la ce trebuia să facă creierul meu în cursul procesării informaţiei. Am continuat această investigaţie timp de cincisprezece ani. Ceea ce voi spune în continuare referitor la modul cum înţelegem acum funcţionarea creierului îţi va părea, dragă cititorule, foarte diferit de conceptul-standard de computer. Pe de altă parte, însă, creierul chiar stochează şi procesează informaţii şi datorită universalităţii computaţiei – concept la care de asemenea mă voi întoarce –, există mai multe paralele între creier şi computer decât ţi s-ar părea la prima vedere.

De fiecare dată când fac ceva – sau mă gândesc la ceva –, Indiferent că este vorba de periatul dinţilor, de păşitul prin bucătărie, de contemplarea unei probleme de afaceri, de exersatul la clape sau de găsirea unei noi idei, reflectez asupra felului cum am reuşit s-o obţin. Mă gândesc şi mai mult la toate lucrurile pe care nu sunt capabil să le fac, deoarece limitările gândirii omeneşti asigură un set la fel de important de indicii. Faptul că mă gândesc atât de mult la gândire ar putea foarte bine să mă încetinească, dar am sperat că asemenea exerciţii de autoreflecţie îmi vor permite să-mi rafinez metodele mintale.

Pentru a ne îmbogăţi conştientizarea modului de

funcţionare a propriului creier, hai să luăm în considerare o serie de experimente mintale.

Încearcă asta: *Spune alfabetul*.

Probabil că îţi aminteşti experimentul din copilărie şi îl poţi face uşor.

În regulă, acum încearcă asta: *Spune alfabetul de la coadă la cap*.

Dacă nu ai studiat alfabetul în această ordine, este posibil să constaţi că îţi este imposibil să realizezi operaţia. Ocazional, câte cineva care a petrecut foarte mult timp într-o sală de clasă de şcoală elementară, unde alfabetul este expus pe pereţi, va putea să facă apel la memoria vizuală şi apoi să-l citească, graţie ei, de la coadă la cap. Însă chiar şi aşa e dificil, deoarece nu ne amintim imagini complete. Recitarea alfabetului în sens invers ar trebui să fie o sarcină simplă, deoarece implică exact aceeaşi informaţie ca şi citirea în sens direct, dar, cu toate acestea, în general suntem incapabili să o facem.

Îţi ţii minte numărul de asigurări sociale? în caz că da, poţi să-l spui invers fără să-l sprii? Dar poezioara de grădiniţă *Mary avea un mieluşel*? Computerele pot face acest lucru în mod curent. Cu toate acestea, noi dăm greş, dacă nu învăţăm în mod expres succesiunea inversă ca pe o serie nouă. Acest fapt ne spune un lucru foarte important despre felul cum este organizată memoria omului.

Desigur, suntem capabili să facem această operaţie cu uşurinţă dacă scriem succesiunea şi apoi o citim de la coadă la cap. Procedând astfel, folosim o tehnologie – limbajul scris – pentru compensarea uneia dintre limitările gândirii noastre neasistate, chit că instrumentul este foarte rudimentar. (A fost cea de-a doua invenţie a noastră, limbajul vorbit fiind prima.) Acesta este motivul pentru care inventăm instrumente – să ne compensăm neajunsurile.

Asta sugerează că memoria omului este secvenţială şi ordonată. Ea poate fi accesată în ordinea în care e reţinută. Nu suntem capabili să inversăm direct secvenţa

unei amintiri.

De asemenea, avem dificultăţi cu declanşarea unei amintiri din mijlocul unei succesiuni. Dacă învăţ să interpretez o piesă muzicală la pian, în general nu o pot începe dintr-un punct arbitrar de la mijloc. Există câteva puncte la care pot sări, deoarece memorarea secvenţială a piesei este organizată în segmente. Însă, dacă încerc să încep în mijlocul unui segment, trebuie să mă întorc la citirea vizuală până când îmi revine memoria secvenţială.

În continuare, încearcă asta: *Rememorează o plimbare pe care ai făcut-o cam cu o zi înainte. Ce îţi aminteşti din cursul ei*?

Acest experiment mintal merge cel mai bine dacă ai făcut o plimbare recentă, cum ar fi astăzi sau ieri. (Poţi să o înlocuieşti cu o plimbare cu maşina sau, practic, cu orice altă activitate în cursul căreia te-ai deplasat pe un teren.)

Este posibil să nu-ţi aminteşti prea multe despre această experienţă. Care a fost cea de-a cincea persoană pe care ai întâlnit-o (fără să incluzi numai oamenii cunoscuţi)? Ai văzut un stejar? O cutie poştală? Ce ai văzut după ce ai dat primul colţ? Dacă ai trecut pe lângă nişte magazine, ce era în a doua vitrină? Probabil că poţi să reconstruieşti răspunsurile la unele dintre aceste întrebări din câteva indicii pe care ţi le aminteşti, dar este posibil să-ţi reaminteşti doar puţine detalii, cu toate că experienţa e foarte recentă.

Dacă faci plimbări în mod regulat, gândeşte-te la prima plimbare pe care ai făcut-o luna trecută (sau la primul drum până la birou de luna trecută, dacă faci naveta). Probabil că nu poţi să-ţi aminteşti deloc respectiva plimbare sau călătorie ca navetist, iar dacă ţi-o aminteşti, fără îndoială că îţi aduci aminte mai puţine detalii legate de ea decât de plimbarea de astăzi.

Voi discuta mai târziu subiectul conştienţei şi voi demonstra că tindem să ne adecvăm conştienţa în funcţie de amintirea evenimentelor. Motivul principal pentru care credem că nu suntem conştienţi când ne aflăm sub efectul

anesteziei este aceia că nu ne amintim nimic din perioada respectivă (cu toate că există excepţii ciudate – şi tulburătoare – în această privinţă). Aşadar, cu privire la plimbarea pe care am făcut-o astăzi de dimineaţă, n-am fost conştient în cea mai mare parte a ei? Este o întrebare raţională, dat fiind faptul că nu-mi amintesc aproape nimic despre ce am văzut şi la ce m-am gândit.

Se întâmplă să existe câteva lucruri pe care mi le amintesc din timpul plimbării de azi dimineaţă. Îmi aduc aminte că m-am gândit la această carte, dar n-aş putea să-ţi spun exact la ce anume m-am gândit. De asemenea, îmi amintesc că am trecut pe lângă o femeie care împingea un cărucior. Îmi aduc aminte că femeia era atrăgătoare şi că bebeluşul era şi el foarte drăgălaş, îmi amintesc două gânduri în legătură cu această experienţă: *Bebeluşul acesta este adorabil, ca nepotul meu nou-născut* şi *Ce o fi înţelegând bebeluşul acesta din împrejurimile vizuale*? Nu-mi amintesc cum era îmbrăcat vreunul dintre ei şi nici ce culoarea avea părul lor. (Soţia mea ţi-ar spune că e ceva obişnuit la mine.) Cu toate că nu sunt capabil să descriu ceva exact cu privire la aspectul lor, am un sentiment inefabil legat de aspectul femeii şi cred că aş putea alege fotografia ei dintre mai multe imagini cu semene de-ale sale. Prin urmare, deşi trebuie să fie o frântură din aspectul ei cea care mi-a rămas în memorie, dacă mă gândesc la femeie, la cărucior şi la copil, nu sunt capabil să vizualizez ansamblul. Nu există nicio înregistrare fotografică sau video a acestui eveniment în mintea mea. Este greu să descriu exact ce anume din mintea mea e legat de această experienţă.

Îmi mai amintesc că am mai trecut pe lângă o femeie cu un cărucior în timpul unei plimbări de acum câteva săptămâni. În acest caz, nu cred că aş mai putea să o recunosc pe femeia aceea într-o fotografie. Amintirea de atunci este mult mai estompată astăzi decât trebuie să fi fost la scurt timp după acea plimbare.

În continuare, gândeşte-te la oamenii pe care i-ai

întâlnit doar o dată sau de două ori. Îi poţi vizualiza clar? Dacă eşti artist în domeniul vizual, e posibil să fi învăţat această abilitate, dar de regulă nu suntem capabili să vizualizăm oameni pe care i-am întâlnit doar ocazional, să-i desenăm sau să-i descriem suficient de bine. Nu am avea însă dificultăţi prea mari în recunoaşterea unei fotografii a lor.

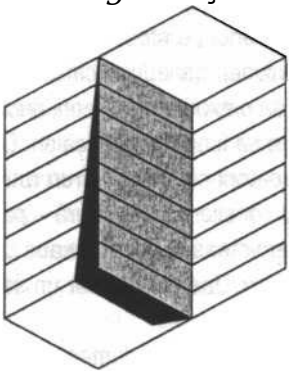
Asta sugerează că în creier nu se stochează imagini, video sau sonore. Amintirile noastre sunt stocate sub forma unor secvenţe de modele. Amintirile care nu sunt accesate se estompează odată cu trecerea timpului. Când artiştii plastici de la poliţie intervievează victima unei infracţiuni, nu o întreabă *Cum erau sprâncenele infractorului*? în schimb, îi arată victimei o serie de imagini de sprâncene şi îi cer să o aleagă pe una dintre ele. Setul corect de sprâncene va genera recunoaşterea modelului stocat în memoria victimei.

Să ne gândim acum la chipurile pe care le cunoşti bine. *îl poţi recunoaşte pe vreunul dintre aceşti oameni*?

Fără îndoială că reuşeşti să identifici personaje familiare, cu toate că sunt parţial acoperite sau distorsionate. Acest fapt constituie unul dintre puncte-cheie ale percepţiei umane: putem recunoaşte un model chiar dacă percepem (vedem, auzim, atingem) numai o parte din el şi chiar dacă modelul conţine distorsiuni. Puterea noastră de recunoaştere este capabilă, după câte s-ar părea, să detecteze trăsături invariabile ale modelului

– caracteristici care supravieţuiesc variaţiilor din lumea reală. Aparentele distorsiuni dintr-o caricatură sau din anumite forme de artă (vezi impresionismul) subliniază modelele unei imagini (persoană, obiect) pe care le recunoaştem, deşi anumite detalii sunt schimbate. Lamea artei se află, de fapt, în avans faţă de lumea ştiinţei în înţelegerea forţei sistemului de percepţie al omului. Folosim aceeaşi abordare şi atunci când recunoaştem o melodie după numai câteva note.

Acum gândeşte-te la această imagine:



Imaginea este ambiguă – colţul indicat de regiunea neagră poate să fie unul interior sau unul exterior. La început, e posibil să-l percepi într-un fel sau în altul, deşi, cu puţin efort, îţi poţi modifica percepţia mergând către o interpretare alternativă. Însă, din momentul în care mintea ta s-a fixat asupra unui înţeles, îţi poate fi foarte dificil să sesizezi cealaltă perspectivă. (Acest lucru se dovedeşte a fi adevărat şi în privinţa perspectivelor intelectuale.) Interpretarea unei imagini de către creierul tău influenţează practic felul în care trăieşti experienţa imaginii respective. Când colţul pare a fi unul interior, creierul tău va interpreta regiunea cenuşie ca fiind o umbră; asta înseamnă că zona nu îţi va părea atât de întunecată ca atunci când interpretezi colţul ca fiind unul exterior.

Prin urmare, trăirea conştientă a experienţei percepţiilor noastre este modificată, de fapt, de interpretările pe care le dăm experienţei respective.

Consideră că vedem ceea ce ne aşteptăm sa.

Sunt încrezător că ai reuşit să completezi propoziţia de mai înainte.

Dacă aş fi scris ultimul cuvânt, ai fi avut nevoie doar de o ocheadă instantanee pentru a-ţi confirma că termenul cu pricina e cel la care te aşteptai.

Asta presupune că anticipăm viitorul în permanenţă şi formulăm ipoteze cu privire la experienţele pe care le vom trăi. Aceste aşteptări influenţează ceea ce percepem de fapt. Anticiparea viitorului este, în principal, motivul fundamental pentru care avem un creier.

Să ne gândim la o experienţă pe care o trăim cu toţii în mod regulat: *o amintire din urmă cu ani de zile ne vine în mod inexplicabil în minte*.

Adeseori este vorba despre amintirea unei persoane sau a unui eveniment la care nu te-ai mai gândit de mult. Este evident că a existat ceva care a provocat amintirea. Firul gândurilor care a făcut posibil acest lucru poate fi evident şi poate fi unul pe care să-l poţi exprima. Alteori, e posibil să fii conştient de succesiunea gândurilor care au dus la revenirea acestei amintiri, dar îţi vine greu să o exprimi. Adeseori, declanşatorul dispare rapid, astfel încât amintirea pare să fi apărut de nicăieri. De multe ori mă confrunt cu asemenea amintiri aleatorii când fac lucruri rutiniere, cum ar fi periajul dinţilor. Uneori pot să fiu conştient de legătură – pasta de dinţi ce cade de pe periuţă poate să-mi amintească de vopseaua care mi-a căzut de pe o pensulă la un curs din liceu. Uneori am doar un sentiment vag al legăturii sau nu am niciun fel de sentiment.

Un fenomen asociat pe care toată lumea îl trăieşte frecvent este încercarea de a-ţi reaminti un nume sau un cuvânt. Procedura pe care o folosim în asemenea împrejurări este să încercăm să vizualizăm declanşatorii capabili să deblocheze amintirea. (De exemplu: *Cine a jucat-o pe Regina Padme în Răzbunarea lui Sâth*? *Să vedem*: *e aceeaşi actriţă care a jucat şi rolul principal într-*

un film recent despre dans, îi zicea Lebăda neagră, a, da, Natalie Portman.) Uneori utilizăm metode de mnemotehnică idiosincretică pentru a ne reaminti. (De exemplu, *E mereu zveltă, nu cât portu, a, da, Portman, Natalie Portman.*) Unele dintre amintirile noastre sunt suficient de solide încât să ne conducă direct de la o întrebare (ca de exemplu *Cine a jucat rolul Reginei Padme7*) la răspuns; alteori, trebuie să verificăm o serie de declanşatori pentru a-l găsi pe cel ce se potriveşte. Este o situaţie care se aseamănă foarte mult cu găsirea link-ului web adecvat. Într-adevăr, e posibil ca amintirile să fie uitate la fel ca o pagină web cu care nu are link-uri nicio altă pagină (cel puţin nicio pagină dintre cele pe care le putem găsi).

În timp ce faci activităţi de rutină – cum ar fi îmbrăcarea cămăşii –, urmăreşte-te efectuându-le şi gândeşte-te la măsura în care urmezi exact aceeaşi succesiune a paşilor de fiecare dată. Din propriile mele observaţii (şi, cum am menţionat, încerc să mă observ permanent), deduc că este posibil să urmezi în cea mai mare măsură aceiaşi paşi de fiecare dată când îndeplineşti o anumită activitate de rutină, cu toate că pot să intervină şi alte module, adăugate. De exemplu, cele mai multe dintre cămăşile mele nu au nevoie de butoni, dar, când e una care are, asta implică o altă serie de sarcini.

Lista paşilor din mintea mea este organizată pe ierarhii. Îndeplinesc o anumită rutină înainte să mă duc la culcare. Primul pas este să-mi periez dinţii. Dar această acţiune este la rândul ei descompusă în serii de paşi mai mici, dintre care primul e să pun pastă pe periuţă. Acest pas, la rândul lui, e alcătuit din paşi şi mai mici, cum ar fi găsirea tubului de pastă, scoaterea capacului şi aşa mai departe. Pasul găsirii tubului are şi el paşii lui, primul dintre aceştia fiind deschiderea dulapului din baie. Acest pas, la rândul său, necesită alţi paşi, primul fiind apucarea exteriorului de la uşa dulapului. Această ramificare continuă, practic, până la o serie de mişcări foarte fine,

astfel încât există literalmente mii de acţiuni mici care formează rutina mea de culcare. Cu toate că mă confrunt poate cu dificultăţi de reamintire a detaliilor plimbării pe care am făcut-o cu numai câteva ore în urmă, nu am nicio greutate în a-mi aminti toţi aceşti nenumăraţi paşi ai pregătirii de culcare – într-o asemenea măsură încât sunt în stare să mă gândesc şi la alte lucruri în timp ce parcurg aceste proceduri. Este important să scot în evidenţă faptul că această listă nu este stocată ca o lungă listă cu mii de paşi – dimpotrivă, fiecare dintre procedurile noastre rutiniere e reţinuta sub forma unei ierarhii elaborate de activităţi asociate.

Acelaşi tip de ierarhie este implicat şi în capacitatea noastră de recunoaştere a obiectelor şi situaţiilor. Recunoaştem chipurile oamenilor pe care îi cunoaştem bine şi recunoaştem de asemenea faptul că aceste chipuri au ochi, un nas, o gură şi aşa mai departe – o ierarhie de modele pe care o folosim atât în percepţiile, cât şi în acţiunile noastre. Utilizarea ierarhiilor ne permite să refolosim modelele. De exemplu, nu trebuie să reînvăţăm conceptul de nas şi gură de fiecare dată când facem cunoştinţă cu un chip nou.

În următorul capitol, vom reuni rezultatele acestor experimente mintale într-o teorie a modului în care ar trebui să funcţioneze neocortexul. Voi demonstra că experimentele dezvăluie atribute esenţiale ale gândirii noastre care sunt uniforme, de la găsirea pastei de dinţi, până la compunerea unei poezii.

Capitolul 3
UN MODEL DE NEOCORTEX:
TEORIA MINŢII BAZATE PE RECUNOAŞTERE

Creierul este un ţesut. E un ţesut complicat, minuţios urzit, cum nu mai e nimic altceva din ce cunoaştem în Univers, dar este format din celule, ca orice ţesut. În mod evident, sunt celule superior specializate, dar ele funcţionează potrivit legilor care guvernează toate celelalte celule. Semnalele lor electrice şi chimice pot fi

detectate, înregistrate şi interpretate, iar compuşii lor chimici pot fi identificaţi; conexiunile care constituie textura creierului pot fi cartografiate. Pe scurt, creierul poate fi studiat exact aşa cum poate fi studiat şi rinichiul.

— David H. Hubel, specialist în neuroştiinţă

Să presupunem că ar exista o maşină a cărei structură ar produce gândirea, senzaţiile şi percepţiile; să ne imaginăm această maşină mărită dar păstrând aceleaşi proporţii, astfel încât să poţi intra în ea ca şi cum ar fi o fabrică. Cu aceste ipoteze, i-ai putea vizita interiorul. Şi ce-ai putea observa acolo? Nimic altceva decât componente care se împing şi se mişcă unele pe altele, niciodată ceva ce ar putea să explice percepţia.

— Gottfried Wilhelm Leibniz

O *ierarhie a modelelor*

Am repetat experimentele şi observaţiile simple descrise în capitolul anterior de mii de ori, într-o multitudine de contexte. Concluziile acestor observaţii mă obligă să explic ce trebuie să facă creierul, la fel cum simplele experimente cu timpul, spaţiul şi masa care s-au efectuat la începutul şi sfârşitul secolului al nouăsprezecelea au impus reflecţiile tânărului maestru Einstein cu privire la funcţionarea Universului. În cadrul discuţiei ce urmează, voi include şi câteva observaţii elementare din domeniul neuroştiinţei, încercând să evit numeroasele detalii care rămân în continuare controversate.

În primul rând, dă-mi voie să explic de ce această secţiune discută exact neocortexul (din latină, cu sensul de *scoarţă nouă*). Ştim cu siguranţă că neocortexul este răspunzător de capacitatea noastră de a face faţă modelelor de informaţie şi că face acest lucru de o manieră ierarhică. Animalele lipsite de neocortex (în esenţă nemamiferele) sunt în mare parte incapabile de înţelegerea ierarhiilor.1 înţelegerea şi supunerea faţă de natura ierarhică înnăscută a realităţii este o caracteristică exclusivă a mamiferelor şi rezultă din faptul că mamiferele

sunt singurele care posedă această recentă structură evolutivă a creierului. Neocortexul este răspunzător de percepţia senzorială, de recunoaşterea a tot ce există, de la obiectele vizuale la conceptele abstracte, la controlul mişcărilor, la judecată, de la orientarea spaţială la gândirea raţională şi la limbaj – adică, în esenţă, ceea ce considerăm că reprezintă „gândirea”.

Neocortexul omului, stratul exterior al creierului, este o structură subţire, în esenţă bidimensională, cu grosimea de aproximativ 2, 5 milimetri (în jur de o zecime de inch). La rozătoare are aproximativ dimensiunile unui timbru poştal şi este neted. O inovaţie evolutivă la primate este aceea că neocortexul s-a cutat minuţios peste restul creierului, având adâncituri profunde, şanţuri şi încreţituri care-i măresc suprafaţa. Datorită acestei configuraţii elaborate, neocortexul reprezintă majoritatea creierului omului, revenindu-i 80 la sută din greutatea acestuia. *Homo sapiens* a căpătat o frunte înaltă pentru a face posibil un neocortex şi mai mare; avem, în special, un lob frontalcare se ocupă de cele mai abstracte modele asociate cu conceptele de nivel înalt.

Această structură subţire este compusă, în esenţă, din şase straturi, numerotate de la I (cel exterior) la VI. Axonii neuronilor din straturile II şi III se proiectează şi în alte regiuni ale cortexului. Axonii (conexiunile de ieşire) din straturile V şi VI sunt conectaţi în principal în afara neocortexului, cu talamusul, trunchiul cerebral şi măduva spinării. Neuronii din stratul IV primesc conexiuni sinaptice (de intrare) de la neuroni care se află în afara neocortexului, în Special de la talamus. Numărul de straturi variază puţin de la o regiune la alta. Stratul IV este foarte subţire în neocortexul motor, deoarece, în această regiune, în cea mai mare parte nu primeşte intrări de ia talamus, de la trunchiul cerebral sau de la măduva spinării. În schimb, în lobul occipital (partea neocortexului răspunzătoare în general de procesarea vizuală) există trei straturi suplimentare care pot fi văzute în stratul IV,

datorită intrărilor considerabile care vin în această regiune, inclusiv de la talamus.

O observaţie de o importanţă fundamentală legată de neocortex priveşte extraordinara uniformitate a structurii sale fundamentale. Aceasta a fost observată pentru prima oară de către un specialist în neuroştiinţă, americanul Vernon Mountcastle (născut în anul 1918). În 1957, Mountcastle a descoperit organizarea neocortexului în formă de coloane. În anul 1978, el a făcut o observaţie la fel de semnificativă pentru neuroştiinţă ca experimentul Michelson-Morley – care, în anul 1887, a infirmat existenţa eterului – pentru fizică. În acel an 1978, Mountcastle a descris organizarea invariabilă a neocortexului, emiţând ipoteza că acesta este format dintr-un singur mecanism care se repetă de foarte multe ori2 şi propunând ca unitate de bază *coloana* corticală. Diferenţele de înălţime dintre anumite straturi din regiunile menţionate anterior sunt pur şi simplu diferenţe ale volumului de interconectare căruia regiunile respective sunt răspunzătoare să îi facă faţă.

Mountcastle a emis şi ipoteza existenţei unor minicoloane în interiorul coloanelor, însă această teorie a fost controversată, din cauză că nu existau demarcaţii vizibile între aceste structuri mult mai mici. Însă experimentele ulterioare au dezvăluit că există, de fapt, unităţi care se repetă în cadrul texturii neuronale a fiecărei coloane. Părerea mea este că unitatea de bază e un recognizator de modele şi că acesta constituie componenta fundamentală a neocortexului. Contrar concepţiei lui Mountcastle referitoare la minicoloane, nu există limite specifice între aceşti recognizori, ei fiind plasaţi unul lângă celălalt de o manieră întrepătrunsă, astfel încât coloana corticală reprezintă pur şi simplu un agregat compus dintr-un mare număr de recognizatori. Aceşti recognizori sunt capabili să se interconecteze între ei pe tot parcursul vieţii. Prin urmare, interconectarea complexă (dintre module) pe care o observăm la nivelul neocortexului nu este perspectficată de codul genetic, el e

creată pentru a reflecta modelele pe care le învăţăm propriu-zis de-a lungul vieţii. Voi prezenta mai detaliat teza enunţată, dar susţin că acesta e modul în care probabil că este organizat neocortexul.

Ar trebui precizat, înainte să analizăm mai departe structura neocortexului, că este important să se realizeze modele ale sistemelor la nivelul adecvat. Deşi chimia se bazează teoretic pe fizică şi poate fi dedusă pe de-a-ntregul din fizică, acest lucru ar fi lipsit de rezultate şi nefezabil în realitate, deoarece chimia şi-a stabilit propriile reguli şi modele. În mod similar, ar trebui să putem deduce din fizică legile termodinamicii, dar, din momentul în care avem un număr suficient de mare de particule ca să le putem numi gaz şi nu nişte particule oarecare, aplicarea ecuaţiilor pentru fizica interacţiunilor fiecărei particule devine imposibil de realizat, în timp ce legile termodinamicii se aplică foarte bine.

La fel, şi biologia îşi are propriile reguli şi modele. O singură celulă a unei insule pancreatice este enorm de complicată, mai ales dacă îi creăm modelul la nivelul moleculelor; realizarea unui model pentru ceea ce face de fapt pancreasul sub raportul reglării nivelului sanguin al insulinei şi sub cel al secreţiei enzimelor digestive este considerabil mai puţin complexă.

Acelaşi principiu se aplică şi nivelurilor de realizare a modelelor şi de înţelegere la nivelul creierului. Fără îndoială, realizarea modelului interacţiunilor la nivel molecular este o parte utilă şi necesară a procesului de retroinginerie a creierului, dar scopul efortului în acest moment este rafinarea modelului pentru a explica felul în care creierul procesează informaţia ca să genereze înţelesuri cognitive.

Omul de ştiinţă american Herbert A. Simon (1916 – 2001), care este creditat drept cofondator al domeniului inteligenţei artificiale, a scris elocvent pe tema înţelegerii sistemelor complexe la nivelul de abstracţie adecvat, în descrierea programului de inteligenţă artificială conceput

de el şi numit EPAM (perceptor şi memorizator elementar), Simon a scris, în anul 1973: „Să presupunem că vrei să înţelegi misteriosul program EPAM pe care îl deţin. Aş putea să-ţi dau două versiuni ale lui. Una ar putea să fie… forma în care l-am scris practic – cu întreaga structură de rutine şi subrutine… Pe de altă parte, aş putea să-ţi dau o versiune în limbaj-maşină a EPAM după ce s-a realizat traducerea completă – după ce a fost uniformizat, ca să spun aşa… Nu cred că trebuie să demonstrez detaliat care dintre aceste două versiuni ar constitui cea mai parcimonioasă, cea mai plină de semnificaţii, cea mai legitimă descriere… Nici măcar nu ţi-o voi propune pe a treia: să nu-ţi dau niciun program, dar în schimb să-ţi pun la dispoziţie ecuaţiile electromagnetice şi condiţiile-limită pe care computerul, considerat sistem fizic, ar trebui să le respecte comportându-se ca EPAM. Acesta ar fi apogeul reducerii şi al incomprehensibilului”.3

Există în jur de o jumătate de milion de coloane corticale într-un creier uman, fiecare ocupând un spaţiu de aproximativ doi milimetri înălţime şi o jumătate de milimetru diametru şi conţinând în jur de 60.000 de neuroni (fapt ce se concretizează într-un total de aproximativ 30 de miliarde de neuroni în neocortex). O estimare grosieră spune că fiecare recognizor de modele dintr-o coloană corticală conţine aproximativ 100 de neuroni, aşa că, în total, în neocortex există ceva de ordinul a 300 de milioane de recognizatori de modele.

Deoarece analizăm modul de funcţionare a acestor recognizatori de modele, dă-mi voie să încep prin a spune că este dificil să ştiu cu precizie de unde să încep. Totul în neocortex se întâmplă simultan, deci nu există niciun început şi niciun sfârşit al proceselor sale. Voi fi obligat adeseori să mă refer la fenomene pe care nu le-am explicat încă, dar la care intenţionez să mă întorc, astfel încât te rog, cititorule, suportă aceste referinţe ulterioare.

Fiinţa umană are o capacitate redusă de procesare logică, dar posedă o capacitate foarte profundă de

recunoaştere a modelelor. Pentru gândirea logică, trebuie să ne folosim neocortexul, care este, în esenţă, un imens recognizator de modele. Nu e un mecanism ideal pentru realizarea transformărilor logice, însă este singurul mijloc pe care îl avem pentru realizarea acestui scop. Să comparăm, de exemplu, modul în care joacă şah un om şi modul de funcţionare a unui program tipic de şah pentru computer. Deep Blue, computerul care l-a învins, în anul 1997, pe Garry Kasparov, campionul mondial uman la şah, era capabil să analizeze implicaţiile logice pentru un număr de 200 de milioane de poziţii pe tabla de joc (reprezentând diferite succesiuni mutare-contramutare) în fiecare secundă. (Apropo, acest lucru poate fi realizat acum de câteva computere personale.) Kasparov a fost întrebat câte poziţii putea să analizeze pe secundă, iar el a răspuns că mai puţin de una. Şi atunci cum de a fost posibil să reuşească să-i facă totuşi faţă lui Deep Blue? Răspunsul constă în capacitatea foarte mare a omului de a recunoaşte modele. Pe de altă parte, trebuie să ne antrenăm această capacitate, motiv pentru care nu oricine poate juca şah la nivel de maestru.

Kasparov învăţase în jur de 100.000 de poziţii pe tabla de joc. Este un număr real – am stabilit că un maestru uman dintr-un anumit domeniu stăpâneşte în jur de 100.000 de calupuri de informaţie. Shakespeare şi-a compus piesele de teatru cu 100.000 de sensuri ale cuvintelor (folosind în jur de 29.000 de cuvinte diferite, însă aplicându-le în numeroase moduri). Sistemele medicale expert care au fost construite pentru a reprezenta cunoştinţele unui medic uman au arătat că un specialist uman din domeniul medical stăpâneşte în jur de 100.000 de concepte din domeniul său. Recunoaşterea unui calup de informaţie din acest depozit nu este simplă, unul şi acelaşi element prezentându-se puţin diferit de fiecare dată când este activat.

Înarmat cu aceste cunoştinţe, Kasparov priveşte tabla de şah, compară modelele pe care le vede cu toate

cele 100.000 de situaţii pe care le stăpâneşte şi face toate aceste 100.000 de comparaţii simultan. Asupra acestui aspect există un consens. Toţi neuronii noştri procesează – analizează – modelele în acelaşi timp. Acest lucru nu înseamnă că toţi *emit impulsuri* simultan (probabil că am cădea la pământ dacă s-ar întâmpla aşa ceva), însă, în timp ce procesează, iau în calcul posibilitatea de a emite impulsuri.

Câte modele poate să stocheze neocortexul? Trebuie să luăm în caicul fenomenul redundanţei. Chipul unei persoane dragi, de exemplu, nu este stocat o singură dată, ci de mii de ori. Unele dintre aceste repetiţii reproduc în mare măsură aceeaşi imagine a chipului, dar majoritatea prezintă diverse perspective ale acestuia, în lumini diferite, cu expresii diferite şi aşa mai departe. Niciunul dintre aceste modele repetate nu este stocat ca imagine propriu-zisă (adică sub forma unor dispuneri bidimensionale de pixeli). Dimpotrivă, e stocat sub forma unor liste de trăsături ale căror elemente proprii modelului sunt ele însele modele. Vom spune în continuare mai exact cum se prezintă aceste ierarhii de caracteristici şi cum sunt ei e organizate.

Dacă pornim de la premisa că toate cunoştinţele fundamentale ale unui expert rezidă în aproximativ 100.000 de calupuri de informaţie (respectiv modele) cu o redundanţă estimată de aproximativ 100 la 1, obţinem un necesar de 10 milioane de modele. Aceste cunoştinţe fundamentale la nivel de expert sunt edificate pe cunoştinţe profesionale mai extinse, deci putem ridica ordinul de mărime al modelelor la circa 30 până la 50 de milioane. Cunoştinţele noastre cotidiene „de bun-simţ” pentru fiinţa umană sunt şi mai numeroase; „descurcăreţii” au nevoie practic de o parte mult mai mare din neocortex decât „erudiţii”. Incluzând şi aceste aspecte, estimarea noastră ajunge ia peste 100 de milioane de modele, cu un factor de redundanţă de circa 100. A se observa că factorul de redundanţă nu este nici pe departe

fix – modelele foarte obişnuite înregistrează uşor un factor de redundanţă de ordinul miilor, în vreme ce un fenomen nou-nouţ poate să aibă factorul de redundanţă mai mic de 10.

După cum voi arăta în continuare, procedurile şi acţiunile noastre conţin de asemenea modele şi sunt stocate tot în regiuni ale neocortexului, astfel încât estimarea pe care o dau eu pentru capacitatea neocortexului uman este de ordinul a câteva sute de milioane de modele. Această evaluare brută se corelează bine cu numărul de recognizatori de modele, pe care îl estimez undeva în jur de 300 de milioane, deci concluzia rezonabilă este că funcţia fiecărui recognizator neocortical de modele constă în procesarea unei iterări (cu alte cuvinte, o copie dintre multiplele copii redundante ale majorităţii modelelor din neocortex) a unui model. Estimarea noastră privitoare la numărul de modele cu care este capabil să opereze creierul omului (inclusiv redundanţa necesară) şi numărul de recognizatori fizici de modele se întâmplă să fie de acelaşi ordin de mărime. Trebuie scos în evidenţă aici faptul că, atunci când mă refer la „procesarea” unui model, vorbesc despre toate lucrurile pe care suntem capabili să le facem cu un model: să-l citim, să-l anticipăm (inclusiv părţi din el), să-l recunoaştem şi să-l implementăm (gândindu-ne mai departe la el sau utilizând un model de mişcare fizică).

Trei sute de milioane de procesoare de modele poate să pără un număr mare şi într-adevăr ar fi suficient pentru a-i permite lui *Homo sapiens* să pună la punct limbajul vorbit şi scris, toate uneltele şi alte creaţii. Aceste invenţii s-au dezvoltat unele după altele, dând naştere unei creşteri exponenţiale a conţinutului de informaţie al tehnologiilor, potrivit descrierii mele din Legea accelerării veniturilor. Nicio altă specie nu a mai realizat acest lucru. După cum am discutat, câteva specii, cum ar fi cimpanzeii, par să posede şi ele o abilitate rudimentară de a înţelege şi a compune limbajul, precum şi unelte primitive. Şi ele au, la

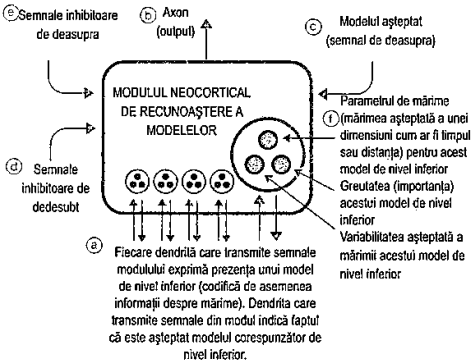
urma urmelor, un neocortex, însă capacităţile lor sunt limitate din cauza dimensiunilor mai mici ale acestuia, în special la nivelul lobilor frontali. Mărimea neocortexului nostru a depăşit un prag care a permis speciei noastre să construiască unelte din ce în ce mai eficiente, inclusiv unele care ne pot permite acum să ne înţelegem propria inteligenţă. În ultimă instanţă, creierul nostru, combinat cu tehnologiile pe care le-a conceput, ne va permite să creăm un neocortex sintetic care va conţine mult peste 300 de milioane de procesatori de modele. De ce nu un miliard? Sau o mie de miliarde?

Structura unui model

Teoria minţii bazate pe recunoaşterea modelelor pe care o voi prezenta aici este întemeiată pe recunoaşterea modelelor de către modulele de recunoaştere a modelelor din neocortex. Aceste modele (ca şi modulele) sunt organizate în ierarhii. Voi prezenta în cele ce urmează rădăcinile intelectuale ale acestei idei, inclusiv propriile mele lucrări în domeniul recunoaşterii ierarhizate a modelelor din anii 1980 şi 1990 şi modelul de neocortex al lui Jeff Hawkins (născut în 1957) şi Dileep George (născut în 1977), de la începutul anilor 2000.

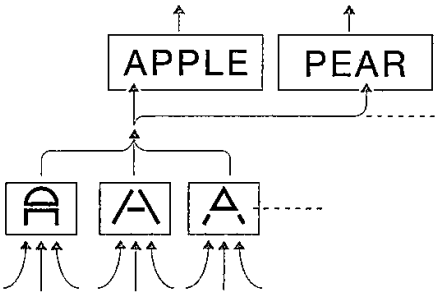
Fiecare model (care este recunoscut de unul dintre cele 300 de milioane de recognizatori de modele estimate că există la nivelul neocortexului)

este compus din trei părţi. Prima parte este inputul, care constă în modelele de nivel inferior ce compun modelul principal. Descrierile fiecăruia dintre aceste modele de nivel inferior trebuie repetată. De exemplu, multe


e Semnale inhibitoare de deasupra
b Axon (output)
c Modelul aşteptat (semnal de deasupra)
MODULUL NEOCORTICAL DE RECUNOAŞTERE A MODELELOR
Parametrul de mărime
f (mărimea aşteptată a unei dimensiuni cum ar fi timpul sau distanţa) pentru acest model de nivel inferior
Greutatea (importanţa) acestui model de nivel inferior
Variabilitatea aşteptată a mărimii acestui model de nivel inferior
d Semnale inhibitoare de dedesubt
a Fiecare dendrită care transmite semnale modulului exprimă prezenţa unui model de nivel inferior (codifică de asemenea informaţii despre mărime). Dendrita care transmite semnale din modul indică faptul că este aşteptat modelul corespunzător de nivel inferior.

dintre modelele cuvintelor includ litera „A”. Niciunul dintre ele nu are nevoie de repetarea descrierii literei „A”, deoarece va folosi aceeaşi descriere. Gândeşte-te la acest lucru ca la un indicator web. Există o pagină web (respectiv un model) pentru litera „A” şi toate paginile web (modelele) pentru cuvintele care conţin „A” vor avea un link la pagina „A” (la modelul „A”), în locul link-urilor pe web, neocortexul foloseşte conexiuni neuronale reale. Există un axon din recognizatorul de modele „A” care se conectează cu mai multe dendrite, câte una pentru fiecare cuvânt care foloseşte „A”. Reţine şi factorul de redundanţă: nu există un singur recognizator de modele pentru litera „A”. Oricare dintre aceşti recognizatori de modele pentru „A” poate să trimită un semnal la recognizatorii de modele care conţin „A”.

A doua parte a fiecărui model este numele acestuia. În lumea limbajului, un model de nivel superior este pur şi simplu cuvântul „apple” („măr”). Cu toate că ne folosim direct neocortexul pentru a înţelege şi procesa fiecare nivel al limbajului, majoritatea modelelor pe care le conţine acesta nu sunt modele de limbaj ca atare. În neocortex, „numele” unui model e pur şi simplu axonul care iese din fiecare procesor de modele; când acest axon emite un impuls, înseamnă că modelul corespunzător a fost recunoscut. Emiterea impulsului de către axon semnifică faptul că recognizatorul strigă numele modelului: „Hei, omule, tocmai am văzut scris cuvântul «apple» («măr»)”.

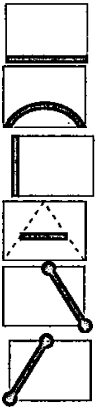


APPLE (măr) PEAR (pară)

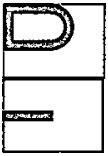
Trei modele redundante (dar oarecum diferite) pentru „A” transmise modelelor superioare care conţin „A”.

A treia şi ultima parte a fiecărui tipar este setul de modele de nivel superior din care face şi el parte. Pentru litera „A”, asta înseamnă toate cuvintele care conţin „A”. Acestea sunt, iarăşi, asemenea link-urilor web. Fiecare model recunoscut la un anumit nivel sesizează următorul nivel că aici e prezentă o parte din modelul de nivel superior. La nivelul neocortexului, aceste link-uri sunt reprezentate de dendrite fizice care se duc la neuronii din fiecare recognizator cortical de modele. Reţine că fiecare neuron poate primi inputuri de la mai multe dendrite, dar că generează un singur output printr-un axon. Acest axon însă poate să transmită la rândul lui semnalul către mai multe dendrite.

Să luăm câteva exemple simple: modelele de pe paginile care urmează reprezintă un mic subset de modele folosite pentru alcătuirea literelor tipărite. Se observă că fiecare nivel reprezintă un model. În acest caz, formele sunt modele, literele sunt modele şi cuvintele sunt de asemenea modele. Fiecare dintre aceste modele are un set de inputuri, un proces al recunoaşterii modelului (bazat pe inputuri care se petrec în modul) şi un output (care este transmis următorului nivel superior de recognizatori de modele).



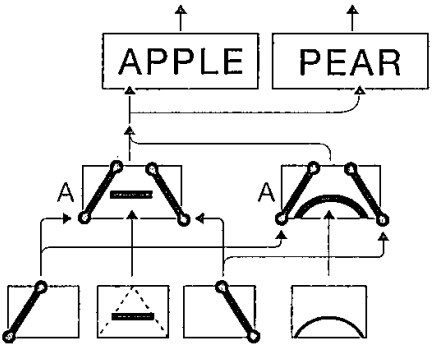
Conexiunea din sud-vest către centru-nord:
Conexiunea sud-est către centru-nord:
Bara orizontală:
Bară verticală de extremă stângă:
Regiunea concavă cu faţa spre sud:
Linia orizontală inferioară:
Linia orizontală superioară:
Linia orizontală medie:



Buclă care formează regiunea superioară:

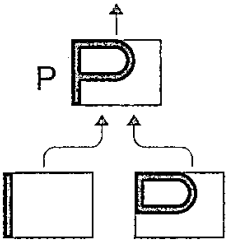
Modelele de mai înainte sunt componente ale următorului nivel superioro categorie denumită litere tipărite (în orice caz, la nivelul neocortexului nu există o asemenea categorie formală; şi, de fapt, nu există categorii formate)

A”.



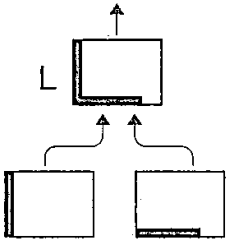
Două modele diferite care constituie fiecare un a „şi două modele diferite la un nivel superior (APPLE” şi. PEAR – măr* şi. pară”) din care face parte. A”.

P”:

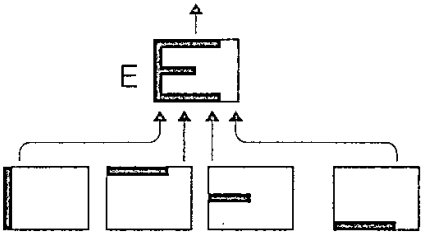


Modele care fac parte din modelul de nivel superior. P

L”:



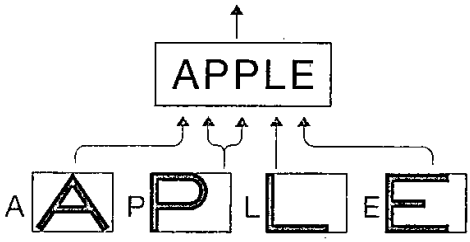
Modele care fac parte din modelul de nivel superior. L”.



Modele care fac parte din modelul de nivel superior. E*

Modelele literelor transmise la un model de nivel superior dintr-o categorie numită „cuvinte”. (Termenul „cuvinte” este categoria noastră de limbaj pentru acest concept, dar neocortexul îl tratează doar ca model.)

„APPLE”:



Într-o parte diferită a cortexului se află o ierarhie comparabilă de recognizatori de modele care procesează imaginile propriu-zise ale obiectelor (şi nu literele tipărite). Dacă priveşti un măr real, recognizătorul de nivel redus detectează marginile curbe şi modelele de culori ale suprafeţei. Acestea fac ca recognizatorul să emită un impuls prin axon şi să spună, de fapt, „Hei, omule, am văzut un măr real”. Alţi recognizatori de modele vor detecta combinaţii de frecvenţe ale sunetelor ce ajung la ei din cortexul auditiv, care poate să-şi excite propriul axon indicând: „Am auzit că s-a rostit cuvântul «măr»”.

Nu uita factorul redundant – nu avem un singur recognizator de model pentru „măr” în fiecare dintre formele lui (scrisă, vorbită, vizuală). Este posibil să existe sute de asemenea recognizatori care să acţioneze, dacă nu chiar mai mulţi. Redundanţa nu numai că măreşte posibilitatea de a recunoaşte cu succes fiecare dintre aceste ipostaze ale mărului, ci se ocupă şi de variaţiile merelor din lumea reală. Pentru merele-obiecte există recognizatori de modele care se ocupă de diversele aspecte ale merelor: culori, nuanţe, forme şi varietăţi.

De asemenea, reţine că ierarhia prezentată anterior este o ierarhie de concepte. Aceşti recognizatori nu sunt plasaţi fizic unii deasupra altora; din cauza subţirimii sale,

neocortexul are, din punct de vedere fizic, înălţimea egală cu a unui singur recognizator. Ierarhia conceptuală este creată de conexiunile dintre recognizatorii individuali de modele.

Un atribut important al Teoriei minţii bazate pe recunoaşterea modelelor îl reprezintă felul în care se fac recunoaşterile în interiorul fiecărui modul de recunoaştere de modele. În modul este stocată o greutate pentru fiecare dendrită de imput care arată cât de important este inputul respectiv pentru recunoaştere. Recognizatorul de modele are un prag de declanşare a impulsului care indică faptul că acest recognizator a recunoscut cu succes modelul pentru care este răspunzător. Nu trebuie să fie prezente toate modelele de imput pentru ca recognizatorul să emită impulsul. El poate emite respectivul impuls şi dacă lipseşte un imput cu greutate redusă, dar e mai puţin probabil să o facă dacă lipseşte un imput cu greutate mare. Când emite impulsul, recognizatorul spune în esenţă: „Modelul de care sunt răspunzător este probabil prezent”.

Recunoaşterea cu succes de către un modul a modelului său depăşeşte simpla numărare a semnalelor de imput care sunt activate (chiar şi o numărare căreia îi conferă greutate importanţa parametrului). Mărimea fiecărui imput contează. Mai există un parametru (pentru fiecare imput) care indică mărimea aşteptată a ihput-ului – şi încă unul care arată cât de variabilă este această mărime. Pentru a înţelege cum funcţionează, să presupunem că avem un recognizator de modele care este răspunzător de recunoaşterea cuvântului „steep” – „abrupt”. Acest cuvânt rostit are patru sunete: [s], [t], [El şi [p]. Fonemul [t] este cunoscut sub numele de „consoană dentală”, ceea ce înseamnă că e creat de limbă, care produce o explozie zgomotoasă când aerul scapă din contactul cu dinţii superiori. În esenţă, este imposibil ca fonemul [t] să se articuleze încet. Fonemul [p] este considerat „consoană plozivă” sau „oral ocluzivă”, ceea ce înseamnă că e produs atunci când tractul vocal este blocat

brusc (de buze în cazul lui [p], astfel încât aerul nu mai trece). Este, de asemenea, obligatoriu, scurt. Vocala [El este produsă de rezonanţa coardelor vocale cu gura deschisă. E considerată o „vocală lungă”, ceea ce înseamnă că persistă o perioadă de timp mult mai lungă decât consoane cum ar fi [t] şi [p]; pe de altă parte, durata poate fi destul de variabilă. Fonemul [s] este cunoscut drept „consoană şuierată” şi e produs de trecerea aerului peste muchiile dinţilor ţinuţi apropiaţi. Durata sa este de obicei mai scurtă decât a unei vocale lungi cum ar fi [El, dar e tot variabilă (cu alte cuvinte, [s] poate fi rostit scurt sau poate fi rostit prelung).

În cercetările noastre dedicate recunoaşterii vorbirii, am descoperit că este necesar ca acest gen de informaţie să fie codificată, pentru recunoaşterea modelelor de vorbire. De exemplu, cuvintele „step” („pas”, „treaptă”) şi „steep” sunt foarte asemănătoare. Cu toate că fonemul [el din „step” şi fonemul [El din „steep” sunt vocale oarecum diferite (prin faptul că au frecvenţe de rezonanţă diferite), nu este de încredere să se facă distincţia între aceste două cuvinte pe baza acestor vocale adeseori confundate între ele. E mult mai de încredere să se ia în considerare observaţia că fonemul [el din cuvântul „step” este relativ scurt în comparaţie cu fonemul [El din cuvântul „steep”.

Putem codifica acest tip de informaţie cu două cifre pentru fiecare imput: mărimea aşteptată şi gradul de variabilitate a acestei mărimi. În exemplul cu „steep”, [t] şi [p] au ambele o durată aşteptată foarte scurtă şi o variabilitate aşteptată redusă (adică nu ne aşteptăm să auzim t-uri şi p-uri lungi). Sunetul [s] are o durată aşteptată mai mică, dar o variabilitate mai mare, deoarece poate fi prelungit. Sunetul [El are atât o durată aşteptată mare, cât şi un grad ridicat de variabilitate.

În exemplele noastre de vorbire, parametrul „mărime” se referă la durată, dar timpul este doar una dintre dimensiunile posibile. De-a lungul cercetărilor noastre de recunoaştere a caracterelor, am constatat că

informaţia spaţială comparabilă era mai importantă pentru recunoaşterea literelor tipărite (de exemplu, punctul de pe litera „i” este aşteptat să fie mult mai mic decât porţiunea de literă de sub punct). La niveluri mult mai înalte de abstractizare, neocortexul poate opera cu modele cu tot felul de continuumuri, cum ar fi nivelurile de atractivitate, ironie, fericire, frustrare şi multe altele. Putem găsi similitudini între continuumuri cât se poate de diferite, cum a făcut Darwin atunci când a asociat mărimea fizică a cantoanelor geologice cu măsura diferenţierii dintre specii.

În cazul unui creier biologic, sursa acestor parametri vine din experienţa proprie a creierului. Nu ne naştem cunoscători ai fonemelor; şi, pe bună dreptate, diferitele limbi au seturi diferite de foneme. Asta presupune faptul că în parametrii învăţaţi ai fiecărui recognizator de modele sunt codificate numeroase exemple ale unui model (deoarece recognizatorul impune mai multe exemple de model pentru a avea siguranţa distribuţiei magnitudinilor inputurilor modelului). În unele sisteme de inteligenţă artificială, aceste tipuri de parametri sunt codificate manual de către experţi (de exemplu de lingvişti, care ne pot preciza duratele aşteptate ale diferitelor foneme, după cum am arătat mai înainte). În cercetările mele, am descoperit că o abordare superioară ar presupune existenţa unui sistem de inteligenţă artificială care să identifice aceşti parametri singur, din datele achiziţionate pe parcursul pregătirii (similar modului în care procedează creierul). Uneori am folosit o abordare hibridă, adică am îmbogăţit sistemul cu intuiţia experţilor umani (pentru setările iniţiale ale parametrilor), iar pe urmă am determinat sistemul de inteligenţă artificială să-şi rafineze automat aceste estimări, folosind un proces de învăţare din exemple reale de vorbire.

Ceea ce face modulul de recunoaştere a modelelor este să computeze probabilitatea (respectiv posibilitatea bazată pe întreaga experienţă anterioară) ca modelul răspunzător de recunoaştere să fie, de fapt, reprezentat pe

moment de inputurile sale active. Fiecare imput ajuns la modul devine activ dacă recognizatorul corespunzător de nivel inferior emite un impuls (ceea ce înseamnă că modelul de nivel inferior a fost recunoscut). Fiecare imput codifică, de asemenea, mărimea observată (pentru o dimensiune adecvată, cum ar fi întinderea temporală, magnitudinea fizică sau un alt continuum), astfel încât mărimea să poată fi comparată de modul, la computarea probabilităţii globale a modelului, cu parametrii de mărime stocaţi pentru fiecare imput.

Cum poate creierul (şi cum poate un sistem de inteligenţă artificială) să computeze probabilitatea globală (pe care modulul este răspunzător să o recunoască) ca modelul să fie prezent, date fiind (1) inputurile (fiecare cu o mărime observată), (2) parametrii stocaţi de mărime (mărimea aşteptată şi variabilitatea mărimii) pentru fiecare imput şi (3) parametrii de importanţă ai fiecărui imput? în anii 1980 şi 1990, împreună cu alţii, am fost pionierul unei metode matematice denumite Modelele Markov Ierarhice Ascunse pentru învăţarea acestor parametri şi pentru folosirea lor ulterioară în vederea recunoaşterii modelelor ierarhice. Am folosit această tehnică în recunoaşterea vorbirii umane şi în înţelegerea limbajului uman. Descriu această abordare mai detaliat în Capitolul 7.

Revenind la fluxul de recunoaştere de la un anumit nivel de recognizatori de modele, în exemplul anterior vedem informaţia curgând în sus pe ierarhia conceptuală, de la caracteristicile elementare ale literelor la cuvinte. Recunoaşterea va continua să se propage de aici mai sus, la fraze şi apoi la structuri şi mai complexe de limbaj. Dacă urcăm alte câteva duzini de niveluri, ajungem la concepte de nivel superior, cum ar fi ironia şi invidia. Deşi toţi recognizatorii de modele funcţionează simultan, este necesar un anumit timp pentru ca recunoaşterea să avanseze în această ierarhie conceptuală. Traversarea fiecărui nivel are nevoie pentru procesare de câteva sutimi

până la câteva zecimi de secundă. Experimentele au dezvăluit că un model de nivel moderat înalt, cum ar fi un chip, are nevoie pentru recunoaştere de cel puţin o zecime de secundă. Este posibil ca procesul să dureze până la o secundă întreagă în cazul în care intervin distorsiuni semnificative. Dacă ar fi secvenţial (cum sunt computerele convenţionale) şi ar realiza recunoaşterea secvenţializată a fiecărui model, creierul ar trebui să ia în calcul toate modelele posibile de nivel inferior înainte să treacă la următorul nivel. Prin urmare, ar fi necesare milioane de cicli doar pentru a se trece prin fiecare nivel. Exact asta se întâmplă atunci când simulăm asemenea procese pe un computer. În orice caz, reţine, cititorule, că un computer procesează de milioane de ori mai rapid decât circuitele noastre biologice.

Un aspect foarte important care trebuie menţionat aici este acela că informaţia se deplasează şi în jos în ierarhia conceptuală, nu numai în sus. Această deplasare în jos este şi mai semnificativă. Dacă, de exemplu, citim de la stânga la dreapta şi am văzut şi am recunoscut deja literele „A”, „P”, „P” şi „L”, recognizatorul pentru „APPLE” anticipează că este posibil să vadă un „E” în următoarea poziţie. Recognizatorul trimite un semnal în jos, la recognizatorul pentru „E”, care spune, de fapt: „Te rog să fii atent, că e o mare probabilitate să vezi modelul tău «E» cât de curând, aşa că stai cu ochii pe el”. Recognizatorul pentru „E” îşi ajustează atunci pragul astfel încât să fie mult mai posibil să recunoască un „E”. Prin urmare, dacă în continuare apare o imagine care e vag asemănătoare cu un „E”, dar poate că este ştearsă, astfel încât n-ar fi fost recunoscută ca „E” în condiţii „normale”, recognizatorul de „E° poate, totuşi, să indice că a văzut într-adevăr un „E”, pentru că se aştepta la asta.

Prin urmare, neocortexul anticipează ce se aşteaptă să întâlnească. Anticiparea viitorului este unul dintre principalele motive pentru care avem neocortex. La nivelul conceptual cel mai înalt, facem în permanenţă predicţii –

cine va intra primul pe uşă, ce e posibil să spună o persoană în continuare, ce ne aşteptăm să vedem după ce trecem de un colţ, rezultatele posibile ale acţiunilor noastre şi aşa mai departe. Aceste predicţii survin în permanenţă, la fiecare nivel al ierarhiei neocortexului. Adeseori identificăm greşit oameni, obiecte şi cuvinte, pentru că pragul nostru de confirmare a unui model aşteptat este prea redus.

Semnalelor pozitive li se adaugă semnale negative sau inhibitoare, care indică faptul că e mai puţin probabil să existe un anumit model. Aceste semnale vin de la nivelurile conceptuale mai joase (de exemplu, recunoaşterea unei mustăţi va inhiba probabilitatea ca persoana pe care o observ că stă la un rând să fie soţia mea), dar şi de la niveluri superioare (de exemplu, ştiu că soţia mea este plecată într-o călătorie, deci persoana de la rând nu poate fi ea). Când un recognizator de modele primeşte un semnal inhibitor, ridică pragul de recunoaştere, dar este în continuare posibil ca modelele să emită impulsuri (deci, dacă persoana de la rând e într-adevăr soţia mea, aş putea s-o recunosc).

Natura datelor care vin la recognizatorul neocortical de modele în continuare, să luăm în considerare. felul cum arată datele pentru un model. Dacă modelul este un chip, datele există în cel puţin două dimensiuni. Nu putem spune că ochii sunt neapărat primii, că sunt urmaţi de nas şi aşa mai departe. Aceeaşi observaţie este valabilă şi în cazul majorităţii sunetelor. O piesă muzicală are cel puţin două dimensiuni. Pot să existe mai multe instrumente şi/sau voci care emit sunete în acelaşi timp. Mai mult, o singură notă a unui instrument complex cum ar fi. pianul este compusă din mai multe frecvenţe. O singură voce umană posedă, simultan, niveluri variabile de energie în zeci de benzi de frecvenţă diferite. Prin urmare, un model de sunet poate fi complex în orice moment şi aceste momente complexe se întind în timp. Imputul tactil este de asemenea bidimensional, deoarece pielea e un organ de

simţ bidimensional şi asemenea modele se pot schimba în funcţie de cea de-a treia dimensiune, timpul.

Aşadar, s-ar părea că inputul pentru un procesor neocortical de modele trebuie să conţină modele bidacă nu chiar tridimensionale. Pe de altă parte, putem observa în structura neocortexului că inputurile de model sunt doar liste unidimensionale. Toate cercetările noastre în domeniul creării sistemelor artificiale de recunoaştere a modelelor (cum ar fi sistemele de recunoaştere a vorbirii şi cele de recunoaştere vizuală) demonstrează că putem (şi am făcut acest lucru) să reprezentăm fenomene bi- şi tridimensionale cu asemenea liste unidimensionale. Voi descrie felul cum funcţionează aceste metode în Capitolul 7, dar deocamdată putem continua cu înţelegerea faptului că inputul pentru fiecare procesor de modele este o listă unidimensională, cu toate că modelul însuşi poate să reflecte în mod inerent mai multe dimensiuni.

Ar trebui să luăm în calcul în acest moment ideea intuitivă că modelele pe care le-am învăţat pentru recunoaştere (de exemplu, un anumit câine sau ideea generală de „câine”, o notă muzicală sau o piesă muzicală) ilustrează exact acelaşi mecanism care stă la baza amintirilor noastre. Amintirile noastre sunt de fapt modele organizate ca liste (în care fiecare componentă, de pe fiecare listă, este un alt model din ierarhia corticală), pe care le-am învăţat şi apoi le recunoaştem atunci când venim în contact cu stimulul adecvat. De fapt, amintirile există în neocortex tocmai pentru a fi recunoscute.

Singura excepţie se află la nivelul conceptual cel mai de jos, la care datele de imput pentru un model reprezintă informaţii senzoriale specifice (de exemplu, date în imagini de la nervul optic). Chiar şi acest cel mai de jos nivel al modelului, pe de altă parte, s-a transformat în mod semnificativ în modele simple înainte să ajungă la cortex. Lista de modele care constituie o amintire are o ordine strictă, aşa că suntem capabili să ne reamintim aceste modele numai în respectiva ordine, de unde şi dificultăţile

pe care le avem în inversarea amintirilor.

Amintirea trebuie să fie declanşată de un alt gând/amintire (ele reprezintă unul şi acelaşi lucru). Putem trăi experienţa mecanismului de declanşare în clipa în care percepem un model. Când am perceput „A”, „P”, „P” şi „L”, modelul „APPLE”„măr” – a anticipat că urma să vedem un „E” şi a declanşat modelul „E” care era aşteptat. Prin urmare, cortexul nostru „gândeşte” că vede un „E” chiar înainte să-l vadă. Dacă această interacţiune din cortexul nostru ne reţine atenţia, ne vom gândi la un „E” chiar înainte să-l vedem sau chiar dacă nu-l vedem deloc. Un mecanism similar ne declanşează amintiri vechi. De obicei este vorba despre un întreg lanţ de asemenea legături. Chiar dacă posedăm un anumit nivel de conştientizare a amintirilor (respectiv modelele) care au declanşat amintirea veche, amintirile (modelele) nu au etichete în limbaj sau imagini. Aceasta este conştientizarea. Acesta este motivul pentru care amintirile vechi pot părea că ne răsar brusc în conştiinţă. Fiind îngropate şi neactivate vreme de ani întregi probabil, au nevoie de un declanşator, tot aşa cum o pagină web are nevoie de un link web pentru a fi activată. Şi, exact aşa cum o pagină web poate ajunge „orfană” din cauză că nicio altă pagină nu conţine un link pentru ea, tot aşa se poate întâmpla şi cu amintirile noastre.

Gândurile ne sunt activate în mare măsură în unul sau două moduri, nedirecţionat şi direcţionat, ambele folosind aceleaşi link-uri corticale. În modul nedirecţionat, lăsăm link-urile să se joace, fără să încercăm să le orientăm spre vreo direcţie anume. Unele forme de meditaţie (cum ar fi meditaţia transcendentală, pe care o practic) se bazează pe lăsarea minţii să facă exact acest lucru. Visele au şi ele aceeaşi proprietate.

În cazul gândirii direcţionate, încercăm să străbatem un proces mult mai ordonat de reapelare a amintirii (o poveste, de exemplu) sau de rezolvare a unei probleme. Iar asta implică parcurgerea listelor din neooortex, însă

agitaţia structurată a gândului nedirecţionat va însoţi şi ea procesul. Conţinutul complet al gândirii noastre este, prin urmare, foarte dezordonat, fenomen pe care James Joyce l-a pus în lumină în romanele sale pe tema „fluxului conştientizării”.

Pe măsură ce gândeşti prin amintirile/poveştile/modelele din viaţa ta, indiferent dacă ele implică întâlnirea întâmplătoare cu o mamă care împinge un cărucior cu un bebeluş în el sau povestea mult mai importantă a felului în care ţi-ai cunoscut soţia/soţul, amintirile rezidă într-o succesiune de modele. Deoarece aceste modele nu sunt etichetate cu cuvinte, sunete, imagini sau video, când încerci să-ţi aminteşti un eveniment semnificativ, în esenţă reconstruieşti în minte modelele, deoarece imaginile propriu-zise nu există.

Dacă ai vrea să „citeşti” gândurile unei persoane şi să arunci o privire asupra a ceea ce se petrece mai exact în neocortexul persoanei respective, ar fi foarte dificil să-i interpretezi amintirile, chiar dacă ai încerca să vizualizezi modelele stocate în neocortex în aşteptarea momentului când vor fi declanşate sau pe cele care au fost declanşate deja şi a căror experienţă este trăită în momentul respectiv sub formă de gânduri active. Ceea ce am „vedea” ar fi activarea simultană a milioane de recognizatori de modele. O sutime de secundă mai târziu, am vedea un set diferit, cu un număr comparabil de recognizatori de modele activaţi. Fiecare asemenea model ar reprezenta o listă a celorlalte modele şi fiecare dintre aceste modele ar fi o listă a celorlalte modele şi aşa mai departe, până am ajunge la modelele elementare, de la nivelul cel mai de jos. Ar fi extrem de dificil să interpretăm semnificaţia modelelor de nivel superior fără *să copiem* în propriul nostru cortex toată informaţia de la fiecare nivel. În acest fel, fiece model din neocortexul nostru este plin de semnificaţii numai în lumina întregii informaţii deplasate la nivelurile de sub el. Mai mult de-atât, alte modele de la acelaşi nivel şi de la niveluri superioare sunt de asemenea

relevante pentru interpretarea unui anumit model, deoarece, asigură contextul. Prin urmare, adevărata „citire a gândurilor” ar necesita nu doar detectarea activărilor axonilor relevanţi din creierul unei persoane, ci şi examinarea întregului neocortex cu toate amintirile sale, pentru înţelegerea acestor activări.

Când trăim experienţa propriilor gânduri şi amintiri, „ştim” ce înseamnă acestea, dar ele nu există ca gânduri şi amintiri uşor explicabile. Dacă vrem să le împărtăşim altora, trebuie să le traducem în limbaj. Şi această sarcină este îndeplinită tot de neocortex, el folosind recognizatori de modele antrenate cu modele care au fost învăţate în scopul folosirii limbajului. Limbajul este el însuşi suficient de ierarhic şi evoluat pentru a profita de natura ierarhică a neocortexului, care, la rândul ei, reflectă natura ierarhică a realităţii. Capacitatea înnăscută a omului de a învăţa structurile ierarhice ale limbajului, despre care a scris Noam Chomsky, reflectă structura neocortexului. Într-o lucrare din anul 2002 al cărei coautor a fost, Chomsky citează atributul „revenirii” ca fiind acela care explică facultatea unică a limbajului la specia umană.4 Revenirea, potrivit lui Chomsky, reprezintă capacitatea de a reuni bucăţi mai mici într-o bucată mare, proces urmat de folosirea bucăţii mari ca parte a unei alte structuri şi continuarea recurentă a acestui proces. În acest fel, suntem capabili să construim structuri elaborate de propoziţii şi paragrafe dintr-un set limitat de cuvinte. Cu toate că Chomsky nu se referă explicit aici la structura creierului, abilitatea pe care o descrie este exact ceea ce face neocortexul.

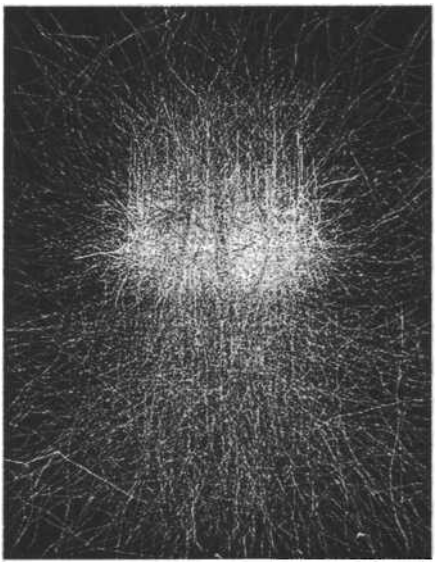
Speciile inferioare de mamifere îşi consumă în mare parte neocortexul cu provocările stilurilor lor particulare de viaţă. Specia umană a căpătat capacităţi suplimentare dezvoltându-şi un cortex substanţial mai mare, cu care poate să se ocupe de limbajul vorbit şi scris. Unii oameni au învăţat această abilitate mai bine decât alţii. Dacă am spune o anumită poveste de multe ori, am începe să

învăţăm practic succesiunea termenilor care descriu povestea ca pe o serie de succesiuni separate. Chiar şi în acest caz, amintirea nu este o succesiune strictă de cuvinte, ci mai curând de structuri de limbaj de care avem nevoie pentru a traduce în succesiuni specifice de cuvinte de fiecare dată când reproducem povestea. Acesta este motivul pentru care spunem povestea puţin diferit de fiecare dată când o împărtăşim (dacă nu învăţăm succesiunea exactă a cuvintelor ca model).

Pentru fiecare dintre aceste descrieri ale proceselor specifice de gândire, mai trebuie să luăm în considerare şi subiectul redundanţei. După cum am menţionat, nu avem nici măcar un model care să reprezinte entităţile importante din viaţa noastră, indiferent dacă aceste entităţi constituie categorii senzoriale, concepte de limbaj sau amintiri ale unor evenimente. Toate evenimentele importante – la toate nivelurile – se repetă de multe ori. Unele dintre aceste recurenţe reprezintă diferite perspective şi puncte de avantaj. Acesta este unul dintre principalele motive pentru care putem recunoaşte un chip familiar din diverse puncte de observare şi în diferite condiţii de iluminare. Fiecare nivel din ierarhie are o redundanţă substanţială, permiţând suficientă variabilitate corespunzătoare acelui concept.

Prin urmare, dacă ar fi să ne imaginăm neocortexul tău când te uiţi la o persoană iubită, am vedea o mulţime de impulsuri nervoase generate de axonii recognizatorilor de modele de la toate nivelurile, pornind de la nivelul de bază al modelelor senzoriale primitive şi ajungând la numeroasele şi diferitele modele ce reprezintă imaginea persoanei iubite. Am sesiza, de asemenea, generarea unui număr foarte mare de impulsuri nervoase reprezentând alte aspecte ale situaţiei, cum ar fi mişcările persoanei respective, cuvintele pe care le spune şi aşa mai departe. Prin urmare, dacă experienţa pare mult mai bogată decât o simplă mişcare sistematică în sus într-o ierarhie de caracteristici, chiar aşa şi este.

Însă mecanismul de bază al ascensiunii într-o ierarhie de recognizatori de modele în care fiecare model conceptual reprezintă un concept mai abstract şi mai integrat rămâne valid. Fluxul de informaţie în direcţie descendentă în ierarhie este şi mai mare, fiecare nivel activat al modelului



Simularea computerizată a generării simultane de impulsuri de către mai mulţi recognizatori de modele din neocortex recunoscut trimiţând predicţii la recognizatorul de modele de la următorul nivel cu privire la ceea ce este posibil să se ivească în continuare. Aparenta bogăţie a experienţei umane este rezultatul, faptului că toate acele sute de milioane de recognizatori de modele din neocortexul nostru îşi evaluează simultan inputurile.

În Capitolul 5, voi discuta despre fluxul de informaţie care vine în neocortex de la pipăit, vedere, auz şi alte organe senzoriale. Aceste inputuri timpurii sunt procesate de regiuni corticale dedicate tipurilor relevante de inputuri senzoriale (cu toate că există o enormă plasticitate a atribuirii acestor regiuni, reflectând uniformitatea funcţiei la nivelul neocortexului). Ierarhia conceptuală continuă peste cele mai înalte concepte în fiecare regiune senzorială a neocortexului. Ariile corticale de asociere integrează

inputurile diferitelor organe de simţ. Când auzim ceva care se aseamănă poate cu vocea soţiei/soţului şi apoi vedem ceva ce poate că-i semnalează prezenţa, nu ne vom angaja într-un proces elaborat de deducţie logică; dimpotrivă, vom percepe instantaneu că e soţul/soţia din combinaţia acestor recogniţii senzoriale. Integrăm toate indiciile senzoriale şi perceptuale relevante – poate chiar şi mirosul parfumului sau al coloniei – într-o percepţie multinivel.

La un nivel conceptual aflat deasupra ariilor corticale senzoriale de asociere, devenim capabili să ne confruntăm – să percepem, să ne reamintim şi să ne gândim la – cu concepte mult mai abstracte. La nivelul cel mai înalt, recunoaştem modele cum ar fi *E amuzant*; *E frumuşică*; *E ironic şi* aşa mai departe. Amintirile noastre includ şi aceste modele de recogniţie abstractă. De exemplu, este posibil să ne reamintim că vorbeam cu cineva care a spus ceva amuzant şi am râs, cu toate că s-ar putea să nu ne mai aducem aminte de gluma propriu-zisă. Succesiunea memoriei acestei amintiri a înregistrat pur şi simplu percepţia umorului, dar nu şi conţinutul exact al comunicării care l-a generat.

În capitolul anterior, am subliniat că ne putem aminti adeseori de un model cu toate că nu-l recunoaştem suficient de bine încât să fim în stare să-l descriem. De exemplu, cred că aş putea să extrag, dintr-un grup de fotografii cu femei, fotografia tinerei cu căruciorul pe care am văzut-o astăzi mai devreme, în ciuda faptului că nu pot. s-o vizualizez de fapt şi nu pot s-o descriu prea amănunţit. În acest caz, amintirea mea legată de ea este o listă cu anumite caracteristici de nivel înalt. Aceste caracteristici nu au ataşate etichete de limbaj sau de imagine şi nu sunt imagini pixelate, astfel încât, deşi pot să mă gândesc la femeie, nu pot să o descriu. Pe de altă parte, dacă mi se dă o fotografie a ei, pot procesa imaginea, fapt ce se concretizează în recunoaşterea aceloraşi caracteristici de nivel înalt care au fost identificate prima oară când am văzut-o. Aş putea să determin, prin urmare, că aceste

caracteristici corespund şi, astfel, aş alege plin de încredere fotografia ei.

Cu toate că am văzut-o pe această femeie o singură dată, la o plimbare, probabil că există deja copii multiple ale modelului ei în neocortexul meu. Pe de altă parte, dacă nu mă gândesc la ea o perioadă de timp dată, aceşti recognizatori de modele se vor atenua: volumul de redundanţă se va diminua până când anumite amintiri se vor stinge. Totodată, acum, după ce mi-am întipărit-o în minte pe această femeie scriind despre ea, probabil că nu o voi uita aşa de uşor.

Asocierea şi invarianţa în capitolul anterior, am discutat despre modul cum putem recunoaşte un model chiar dacă acesta nu e prezent integral şi, de asemenea, chiar dacă este distorsionat. Prima capacitate se numeşte autoasociere: capacitatea de asociere a modelului cu o parte din el însuşi. Structura fiecărui recognizator de modele încorporează în mod inerent această capacitate.

Pe măsură ce fiecare imput de la un recognizator de model inferior urcă la unul de nivel superior, conexiunea poate să aibă o „greutate” care să indice cât de important este acel element anume din model. Astfel, elementele cele mai semnificative ale unui model cântăresc mai mult atunci când se apreciază că modelul respectiv ar trebui să fie considerat „recunoscut”. Este foarte posibil ca barba lui Lincoln, perciunii lui Elvis şi celebrul gest al lui Einstein de a scoate limba să cântărească greu în modelele pe care le ştim, deoarece aparţin acestor personaje reprezentative. Recognizatorul de model computează o probabilitate care ia în calcul parametrii de importanţă. Astfel, probabilitatea de ansamblu este mai redusă dacă lipseşte un element sau mai multe, deşi pragul de recunoaştere poate să fie, totuşi, satisfăcut. După cum am subliniat, computaţia probabilităţii de ansamblu (că este prezent modelul) e mai complicată decât o simplă sumă cântărită în ideea că parametrii de mărime trebuie luaţi şi ei în calcul.

Dacă recognizatorul de model a primit semnal de la

un recognizator de nivel superior că modelul lui este „aşteptat”, atunci pragul este redus corespunzător (respectiv, e făcut mai uşor de accesat). Reversul: un asemenea semnal poate să se adauge pur şi simplu la totalul inputurilor cântărite, compensând prin aceasta un element-lipsă. Asta se întâmplă la toate nivelurile, astfel încât un model cum ar fi un chip, aflat la mai multe niveluri deasupra celui mai de jos, poate fi recunoscut până şi în condiţiile lipsei mai multor caracteristici.

Capacitatea de recunoaştere a modelelor chiar şi în condiţiile transformării unor aspecte ale acestora se numeşte Invarianţa caracteristicilor şi se realizează în patru feluri. În primul rând, există transformări globale care sunt realizate înainte ca neocortexul să primească datele senzoriale. Vom discuta drumul datelor senzoriale de la ochi, urechi şi piele în secţiunea *Calea senzorială*, la pagina 93.

A doua metodă profită de redundanţa memoriei corticale de modele. Mai ales în cazul subiectelor importante, am acumulat multe perspective diferite şi puncte de observare pentru fiecare model. Prin urmare, numeroase variaţii sunt stocate şi procesate separat.

Cea de-a treia şi cea mai eficientă metodă ţine de abilitatea combinării a doua liste. Una dintre liste poate să conţină un set de transformări despre care ştim că se pot aplica unei anumite categorii de modele; cortexul aplică exact această listă de posibile schimbări la un alt model. În acest fel înţelegem fenomene cum ar fi metaforele şi comparaţiile.

De exemplu, am învăţat că anumite foneme (sunetele de bază ale unei limbi) pot să lipsească din limba vorbită (de exemplu, „goin”). Dacă învăţăm pe urmă un alt cuvânt vorbit (de exemplu „driving”), probabil că vom putea să recunoaştem acest cuvânt şi când lipseşte unul dintre fonemele sale, în ciuda faptului că nu am mai avut experienţa respectivului cuvânt în această formă, deoarece ne-am familiarizat cu fenomenul general al omiterii

anumitor foneme. Alt exemplu: putem învăţa că unui anumit artist îi place să scoată în evidenţă (făcându-le mai mari) anumite elemente ale feţei, cum ar fi nasul. Putem atunci identifica un chip cu care suntem familiarizaţi şi căruia i-a fost aplicată această modificare chiar dacă nu am mai văzut modificarea respectivă aplicată acelui chip.

Anumite modificări artistice scot în evidenţă chiar caracteristicile care sunt recunoscute de neocortexul nostru bazat pe recunoaşterea modelelor. După cum am menţionat, exact acesta este fundamentul caricaturii.

A patra metodă rezultă din parametrii de mărime care-i permit unui singur modul să codifice mai multe aspecte ale unui model. De exemplu, am auzit de multe ori cuvântul „steep" – „abrupt". Un anumit modul de recunoaştere a modelului care identifică acest cuvânt rostit poate codifica multiple exemple indicând că durata lui [El are o înaltă variabilitate de aşteptări. Dacă toate modulele alocate cuvintelor care conţin [El împărtăşesc un fenomen similar, această variabilitate poate fi codificată chiar în modelele lui [El. Pe de altă parte, diferitele cuvinte care conţin [El (sau multe alte foneme) pot să aibă variabilităţi aşteptate diferite. De exemplu, cuvântul „peak" – „vârf1 – e posibil să nu conţină fonemul [Ei la fel de evident cum o face cuvântul „steep".

Învăţarea t

Oare nu ne creăm noi înşine succesorii la supremaţia pe Pământ, contribuind zilnic la frumuseţea şi delicateţea organizării lor, dându-le zilnic mai multă îndemânare şi punându-le la dispoziţie din ce în ce mai mult din acea forţă automată şi autoreglată care va fi mai bună decât orice intelect?

— Samuel Butler, 1871

Principalele activităţi ale creierului constau în a se modifica pe el însuşi.

— Marvin Minsky, Societatea Minţii

Până acum am examinat modul în care recunoaştem (senzorial şi perceptiv) modele şi ne amintim succesiuni de

modele (amintirile legate de lucruri, oameni şi evenimente). Pe de altă parte, nu ne-am născut cu un neocortex plin cu toate aceste modele. Neocortexul nostru este un teritoriu virgin în momentul când ia naştere creierul. Are capacitatea de a învăţa şi prin urmare de a crea conexiuni între recognizatorii de modele, însă dobândeşte aceste conexiuni prin experienţă.

Procesul de învăţare începe chiar înainte de a ne naşte, producându-se concomitent cu procesul biologic de creştere propriu-zisă a creierului. Fătul are deja creier la vârsta de o lună, cu toate că, în esenţă, acesta este un creier reptilian, fătul parcurgând de fapt, în uter, la mare viteză, evoluţia biologică. Creierul devine în mod distinct un creier uman cu un neocortex uman înainte să ajungă în al treilea trimestru al sarcinii.

În acest moment, fătul trăieşte experienţe, iar neocortexul învaţă. Aude sunete, mai ales bătăile de inimă ale. mamei, acesta fiind, foarte posibil, un motiv pentru care calităţile ritmice ale muzicii sunt universale în cultura omului. Absolut toate civilizaţiile descoperite vreodată au avut ca parte a culturii lor muzica, spre deosebire de alte forme ale artei, cum ar fi arta plastică. De asemenea, este şi motivul pentru care ritmul muzicii este comparabil cu ritmul bătăilor inimii. În mod evident, ritmurile muzicii variază – altminteri, muzica nu ne-ar mai reţine interesul –, dar bătăile inimii variază şi ele. Cele excesiv de regulate reprezintă, de fapt, un simptom de boală cardiacă. Ochii fătului sunt parţial deschişi la douăzeci şi şase de săptămâni de la concepere şi sunt deschişi complet aproape tot timpul înainte de a douăzeci şi opta săptămână de la concepere. Poate că nu sunt prea multe de văzut în uter, dar există modele de întuneric şi lumină pe care neocortexul începe să le proceseze.

Aşadar, deşi nou-născutul beneficiază în uter de o mică experienţă, aceasta este clar limitată. De asemenea, neocortexul poate să înveţe de la creierul vechi (un subiect pe care îl discut în Capitolul 5), dar, în general, la naştere,

copilul are multe de învăţat – aproape totul, de la sunetele şi formele primitive de bază, până la metafore şi sarcasm.

Învăţarea este fundamentală pentru inteligenţa umană. Dacă ar trebui să realizăm un model perfect ca să simulăm neocortexul omului (aşa cum încearcă să facă Proiectul Blue Brain) şi toate celelalte regiuni de care creierul are nevoie pentru a funcţiona (de pildă hipocampul şi talamusul), acest neocortex artificial nu ar putea face prea multe – la fel cum nou-născutul nu poate face prea multe (în afară de a fi drăgălaş, ceea ce reprezintă în mod cert o adaptare-cheie pentru supravieţuire).

Învăţarea şi recunoaşterea au loc simultan. Începem să învăţăm imediat şi, de îndată ce am învăţat un model, începem să-l recunoaştem. Neocortexul încearcă permanent şi continuu să găsească logica inputului care îi este pus la dispoziţie. Dacă un anumit nivel se dovedeşte incapabil să proceseze complet şi să recunoască un model, acesta este trimis la următorul nivel superior. Dacă niciunul dintre niveluri nu reuşeşte să recunoască modelul, acesta este considerat nou. Clasificarea unui model ca fiind nou nu înseamnă în mod necesar că toate aspectele sale sunt noi. Dacă ne uităm la picturile unui anumit artist şi vedem o pisică având o trompă de elefant, putem identifica fiecare dintre trăsăturile caracteristice, dar vom observa că acest model combinat este ceva nou şi că e posibil să ni-l reamintim. Nivelurile conceptuale superioare ale neocortexului, care înţeleg contextul – de exemplu, circumstanţa că această pictură este un exemplu al operei unui anumit artist şi că participăm la deschiderea unei expoziţii de tablouri noi ale respectivului artist –, vor constata combinaţia neobişnuită de modele din chipul pisicii-elefant, dar în acelaşi timp vor include aceste detalii contextuale ca modele suplimentare de amintire.

Amintirile noi, cum ar fi chipul pisicii-elefant, sunt stocate într-un recognizator de modele disponibil. Hipocampul joacă un rol în acest proces şi vom discuta ce

se ştie despre mecanismele biologice propriu-zise în capitolul următor. Pentru scopurile modelului nostru de neocortex, este suficient să spun că modelele care nu sunt recunoscute sunt stocate ca modele noi şi sunt conectate în mod adecvat la modelele de nivel inferior care le formează. Chipul pisicii-elefant, de exemplu, va fi stocat în mai multe moduri. Noul aranjament de componente ale feţei va fi stocat ca o amintire contextuală care îl include şi pe artist, şi situaţia ori poate faptul că am râs când l-am văzut prima oară.

Amintirile care sunt recunoscute cu succes pot să se concretizeze şi în crearea unui nou model de atingere a unei redundanţe mai mari. Dacă modelele nu sunt recunoscute perfect, este posibil să fie stocate ca şi cum ar reflecta o perspectivă diferită a subiectului care a fost recunoscut.

Şi atunci, care e metoda general valabilă pentru a stabili ce modele sunt stocate? în termeni matematici, problema poate fi formulată după cum urmează: Folosind limitele disponibile de stocare a modelelor, cum putem reprezenta optim modelele de imput cu care ne-am confruntat până acum? Cu toate că este logic să se accepte un anumit nivel de redundanţă, nu ar fi practic să fie umplută întreaga arie de stocare (respectiv întregul neocortex) cu modele repetate, de vreme ce asta nu ar permite o suficientă diversitate a modelelor. Un model cum ar fi cel al fonemului [El în cuvintele rostite este unul de care ne-am lovit de nenumărate ori. E un model simplu de frecvenţe sonore şi fără îndoială că se bucură de o redundanţă semnificativă în neocortexui nostru. Am putea să ne umplem întregul neocortex cu modele repetate ale fonemului [El. Există, totuşi, o limită pentru redundanţa utilă şi un model obişnuit cum este acesta în mod clar a atins-o.

Există o soluţie matematică la această problemă de optimizare: programarea liniară. Aceasta rezolvă alocarea cea mai bună cu putinţă a resurselor limitate (în cazul de

faţă, un număr limitat de recognizatori de modele) care ar reprezenta toate cazurile cu care s-a pregătit sistemul. Programarea liniară este concepută pentru sisteme cu inputuri unidimensionale, acesta fiind un alt motiv pentru care e optim să reprezentăm inputul fiecărui modul de recunoaştere de model ca pe un şir liniar de inputuri. Putem folosi această abordare matematică în cazul unui sistem software şi, cu toate că creierul real este limitat şi mai mult de conexiunile fizice pe care le are la dispoziţie şi pe care le poate adapta între recognizatorii de modele, metoda este totuşi similară.

O importantă implicaţie a acestei soluţii optime este că experienţele rutiniere sunt recunoscute, dar nu se concretizează în realizarea unei amintiri definitive. În privinţa plimbării mele, m-am confruntat cu milioane de modele la fiecare nivel, de la muchiile şi nuanţele vizuale de bază, la obiecte cum ar fi stâlpii de iluminat, cutiile poştale, oamenii, animalele şi plantele pe lângă care am trecut pe drum. Aproape nimic de pe parcursul acestei experienţe nu a fost ceva unic şi modelele pe care le-am recunoscut îşi atinseseră de mult nivelul optim de redundanţă. Rezultatul este că nu-mi amintesc aproape nimic din cursul acestei plimbări. Cele câteva detalii de care îmi aduc aminte e posibil să fie înlocuite cu noi modele înainte să fac câteva zeci de noi plimbări, exceptând faptul că am memorat această plimbare anume scriind despre ea.

Un aspect important care se aplică atât neocortexului nostru biologic, cât şi tentativelor de a-l reproduce este acela că e dificil să înveţi simultan mai multe niveluri conceptuale. În esenţă, putem învăţa unui, maximum două asemenea niveluri în acelaşi timp. Odată ce învăţarea devine relativ stabilă, putem trece la asimilarea următorului nivel. Putem continua să operăm reglajul fin al învăţării pe nivelurile inferioare, dar învăţarea se concentrează pe următorul nivel de abstractizare. Acest lucru este adevărat atât la începutul vieţii, când nou-

născutul se luptă cu formele de bază, cât şi mai târziu în timpul vieţii, când ne luptăm să învăţăm o altă materie, pe rând, pe câte un nivel de complexitate odată. Descoperim acelaşi fenomen la maşinile care imită neocortexul. Pe de altă parte, dacă suntem confruntaţi cu un material din ce în ce mai abstract, la un moment dat maşina este capabilă să înveţe la fel ca omul (deşi deocamdată nu pe tot atât de multe niveluri conceptuale).

Outputul unui model poate face legătura cu un model de la un nivel inferior sau chiar cu modelul propriu-zis, conferindu-i creierului uman puternica abilitate recursivă. Elementul unui model poate fi un punct de decizie bazat pe un alt model. Acest fapt este deosebit de util pentru listele care compun acţiuni – de exemplu, cumpărarea unui nou tub de pastă de dinţi dacă actualul este gol. Aceste condiţionări există la toate nivelurile. După cum ştie oricine a încercat să programeze o procedură, condiţionalele sunt vitale pentru descrierea cursului unei acţiuni.

Limbajul gândirii

Visul funcţionează ca o supapă de siguranţă pentru creierul supraîncărcat.

— Sigmund Freud, *interpretarea viselor*, 1911

Creierul: un aparat cu care gândim că gândim.

— Ambrose Bierce, *Dicţionarul diavolului*

Ca să rezum ceea ce am aflat până acum despre modul de funcţionare a neocortexului, te rog să te uiţi peste diagrama modulului neocortical de recunoaştere a modelelor de la pagina 45.

a. Dendritele pătrund în modulul care reprezintă modelul. Deşi modelele pot părea că au o natură bisau tridimensională, sunt reprezentate de o succesiune de semnale unidimensionale. Modelul trebuie să fie prezent în această ordine (secvenţială) pentru ca recognizatorul de modele să fie capabil să-l recunoască. Fiecare dendrită este conectată în primul rând la unul sau la mai mulţi axoni ai recognizatorilor de modele, la nivelul conceptual

cel mai redus care a recunoscut un model de nivel inferior ce constituie o parte din acest model. Pentru fiecare dintre modelele din imput, pot exista mai mulţi recognizatori de modele de nivel inferior, care pot genera semnalul că modelul de nivel inferior a fost recunoscut. Pragul necesar pentru recunoaşterea modelului poate fi asigurat chiar dacă râu toate inputurile au transmis semnale. Modulul computează probabilitatea de a fi prezent modelul de care este răspunzător. Această computaţie ia în calcul parametrii „importanţă” şi „mărime” (vezi [I] în continuare).

Reţine că unele dendrite transmit semnale spre modul, iar altele dinspre modul. Dacă toate dendritele de imput ale acestui recognizator de modele semnalează că modelele lor de nivel redus.au fost recunoscute, cu excepţia a unui sau două, recognizatorul va transmite un semnal în jos la recognizatoru! (-oarele) de modele, recunoscând modelele de nivel inferior care nu au fost încă recunoscute şi indicând faptul că există o mare probabilitate ca acele modele să fie recunoscute curând, aşa că recognizatorul (rii) de nivel redus ar trebui să îl aştepte.

b. Când acest recognizator de modele îşi recunoaşte modelul (pe baza activării tuturor sau a majorităţii semnalelor de imput de la dendrite), se activează axonul (output) acestui recognizator de modele. La rândul lui, acest axon poate să se conecteze la o întreagă reţea de dendrite conectate la recognizatori de modele de nivel superior, la care este transmis acest model ca imput. Acest semnal transmite magnitudinea informaţiei astfel încât recognizatorii de modele de la următorul nivel superior conceptual să-l poată lua în considerare.

c. Dacă un recognizator de modele de nivel superior primeşte un semnal pozitiv de la toate sau de la majoritatea modelelor constituente cu excepţia celui reprezentat de un alt recognizator de modele, recognizatorul de nivel superior poate să transmită un

semnal la respectivul recognizator, prin care să-l anunţe că modelul său este aşteptat. Un asemenea semnal ar determina recognizatorul de model să îşi reducă pragul, ceea ce înseamnă că ar fi mai posibil să transmită la axonul lui un semnal care să indice că modelul este recunoscut chiar dacă o parte dintre inputurile sale lipsesc sau sunt neclare.

d. Semnalele inhibitoare de dedesubt ar reduce posibilitatea ca acest recognizator de modele să-şi recunoască modelul. Asta poate rezulta din identificarea modelelor de nivel inferior care sunt inconsistente cu modelul asociat acestui recognizator de modele (de exemplu, recunoaşterea unei mustăţi de către un recognizator de nivel inferior ar face mai puţin probabil ca imaginea respectivă să fie „soţia mea”).

e. Semnalele inhibitorii de deasupra ar face de asemenea mai puţin probabil faptul ca acest recognizator de modele să îşi recunoască modelul. Această situaţie poate să rezulte dintr-un context de nivel superior care este inconsecvent cu modelul asociat acestui recognizator.

f. Pentru fiecare imput, sunt stocaţi parametri legaţi de importanţă, de mărimea aşteptată şi de variabilitatea aşteptată a mărimii. Pe baza tuturor acestor parametri şi a semnalelor curente care indică inputurile prezente şi magnitudinile lor, modulul computează o probabilitate de ansamblu ca modelul să fie prezent. Un mijloc matematic optim de asigurare a acestui rezultat îl reprezintă *Modelele Markov Ascunse*. Când asemenea modele sunt organizate într-o ierarhie (aşa cum se întâmplă în neocortex sau în tentativele de simulare a neocortexului), le numim Modele Markov Ascunse Ierarhic.

Modelele generate de neocortex generează alte modele. Modelele parţial complete transmit semnale în jos în ierarhia conceptuală; modelele complete transmit semnale în sus în respectiva ierarhie. Aceste modele neocorticale constituie limbajul gândurilor. La fel ca limbajul, şi ele sunt ierarhizate, dar nu sunt un limbaj ca

atare. Gândurile noastre nu sunt concepute în principal în elemente ale limbajului, cu toate că, de vreme ce limbajul există în neocortexul nostru şi ca ierarhii de modele, putem avea gânduri bazate pe limbaj. Însă, în cea mai mare parte a lor, gândurile sunt reprezentate în aceste modele neocorticale.

Aşa cum am discutat mai înainte, chiar dacă reuşeam să detectăm activările de model din neocortexul cuiva, tot am fi avut prea puţină idee despre ce înseamnă acele activări de model în lipsa accesului la întreaga ierarhie de modele de deasupra şi de dedesubtul fiecărui model activat. Acest fapt ar presupune, într-o măsură destul de mare, accesul la întregul neocortex al persoanei respective. Ne este dificil să înţelegem conţinutul propriilor gânduri, iar înţelegerea gândurilor altcuiva necesită stăpânirea unui alt neocortex, străin. Desigur, încă nu avem acces la neocortexul altei persoane; trebuie, prin urmare, să ne bazăm pe încercările acelei persoane de a-şi exprima gândurile prin limbaj (precum şi prin alte mijloace, cum ar fi gesturile). Abilitatea incompletă a omului de a-şi îndeplini aceste necesităţi de comunicare mai adaugă un strat de complexitate – nu e de mirare că ne înţelegem unii pe alţii atât de greşit.

Avem două moduri de gândire. Unul îl reprezintă gândirea indirectă, în cadrul căreia gândurile se generează unele pe altele în mod non-logic. Când ne confruntăm cu rememorarea bruscă a unei amintiri din urmă cu ani sau decenii în timp ce facem altceva, de pildă greblăm frunze sau mergem pe stradă, experienţa este rememorată – aidoma tuturor amintirilor – ca o succesiune de modele. Nu vizualizăm imediat scena decât dacă putem reapela o sumedenie de amintiri care ne permit să sintetizăm o amintire mai robustă. Dacă izbutim să vizualizăm o scenă în acest mod, în esenţă o recreăm în mintea noastră din indicii, în momentul reamintirii; amintirea propriu-zisă nu este stocată sub formă de imagini sau de vizualizări. După cum am menţionat anterior, generatorii care determină

acest gând să ne răsară în minte pot să fie sau pot să nu fie evidenţi. E posibil ca succesiunea gândurilor relevante să fi fost dată uitării pe loc. Chiar dacă ne reamintim o scenă, aceasta va reprezenta o succesiune neliniară şi ocolită de asocieri.

Al doilea mod de a gândi este gândirea direcţionată, la care apelăm atunci când încercăm să rezolvăm o problemă sau să formulăm un răspuns organizat. De exemplu, este posibil să repetăm în gând ceea ce intenţionăm să-i spunem cuiva sau să formulăm un pasaj pe care vrem să-l scriem (chiar şi într-o carte mintală). Când ne gândim la sarcini de acest gen, deja le-am descompus pe rând într-o ierarhie de subsarcini. Scrierea unei cărţi, de exemplu, presupune scrierea unor capitole; fiecare capitol conţine secţiuni; fiecare secţiune conţine paragrafe; fiecare paragraf conţine propoziţii care exprimă idei; fiecare idee are propria configurare de elemente; fiecare element şi fiecare relaţie dintre elemente este o idee şi ea trebuie articulată – şi aşa mai departe. Totodată, structurile noastre neocorticale au învăţat anumite reguli care ar trebui respectate. Dacă sarcina este scrisul, atunci s-ar impune să evităm repetiţia inutilă; ar trebui să încercăm să ne asigurăm că cititorul poate urmări ce citeşte; ar trebui să ne străduim să respectăm reguli de gramatică şi stil şi aşa mai departe. Scriitorul trebuie, prin urmare, să-şi construiască în minte un model al cititorului, iar o asemenea construcţie este de asemenea ierarhică. Prin gândirea direcţionată, străbatem listele în neocortex, fiecare dintre ele extinzându-se în ierarhii vaste de subliste, fiecare cu propriile considerente. Nu uita că elementele de pe o listă aflată într-un model neocortical pot include condiţionări, aşa încât gândurile şi acţiunile noastre ulterioare vor depinde de evaluările făcute pe măsură ce străbatem procesul.

Mai mult de-atât, fiecare dintre gândurile direcţionate va genera ierarhii de gânduri nedirecţionate. Un continuu iureş de reflectare ne slujeşte atât

experienţele senzoriale, cât şi încercările de gândire direcţionată. Experienţa noastră mentală propriu-zisă este complexă şi amestecată, fiind compusă din furtuni cu fulgere de modele generate, care se schimbă de aproximativ o sută de ori pe secundă.

Limbajul viselor

Visele sunt exemple de gânduri nedirecţionate. Au o anumită logică, deoarece fenomenul de generare a unui gând de către alt gând se bazează pe conexiuni reale între modele la nivelul neocortexului nostru. În măsura în care visul nu este logic, încercăm să-l reparăm prin capacitatea noastră de a născoci. Aşa cum voi demonstra în Capitolul 9, pacienţii cu creierul secţionat (al căror corp calos, care conectează cele două emisfere ale creierului, este secţionat sau lezat) născocesc (inventează) explicaţii cu ajutorul creierului stâng – care controlează centrul vorbirii – pentru a lămuri ceea ce tocmai a făcut creierul drept cu un input la care creierul stâng nu a avut acces. Născocim în permanenţă, explicând rezultatul evenimentelor. Dacă vrei un bun exemplu în acest sens, caută comentariile cotidiene pe tema mişcărilor de pe pieţele financiare. Indiferent cum performează pieţele, este întotdeauna posibil să se născocească o explicaţie bună pentru motivul care a dus la efectul respectiv; iar asemenea comentarii post factum există din abundenţă. Desigur, dacă numeroşii comentatori ar înţelege cu adevărat pieţele, nu ar trebui să-şi irosească timpul făcând comentarii.

Actul născocirii este realizat, desigur, şi de neocortex, care se pricepe să inventeze poveşti şi explicaţii capabile să satisfacă anumite constrângeri. Facem acest lucru de fiecare dată când spunem din nou o poveste. Includem detalii care poate că nu ne sunt accesibile sau pe care poate că le-am uitat, în aşa fel încât povestea să fie mai logică. Acesta este motivul pentru care poveştile se schimbă cu timpul pe măsură ce sunt spuse de nenumărate ori de noi povestitori, care poate că au interese diferite. Pe de altă parte, la fel cum limbajul vorbit a dus la limbajul

scris, tot aşa am avut la dispoziţie o tehnologie care a putut înregistra versiunea definitivă a unei poveşti şi a putut împiedica acest gen de derivă.

Conţinutul propriu-zis al unui vis, în măsura în care ni-l mai reamintim, este iarăşi o succesiune de modele. Aceste modele reprezintă constrângerile dintr-o povestire; pe urmă născocim o poveste care să corespundă acestor constrângeri. Versiunea visului pe care o povestim (fie şi nouă înşine, muteşte) este chiar această născocire. Când povestim un vis, declanşăm cascade de modele care umplu visul real, aşa cum l-am trăit de fapt.

Între gândurile din vise şi gândirea din Starea de veghe există o deosebire-cheie. Una dintre lecţiile pe care le-am învăţat în viaţă e că anumite acţiuni, chiar şi gânduri, nu sunt permise în lumea reală. De exemplu, am învăţat că nu putem să ne îndeplinim imediat visele. Există reguli împotriva smulgerii banilor-gheaţă din sertarul casierului de la magazin şi constrângeri în interacţiunile cu o persoană de care putem fi atraşi fizic. De asemenea, am învăţat că anumite gânduri nu sunt permise pentru că sunt interzise de cultură. Pe măsură ce ne însuşim deprinderi profesionale, învăţăm şi modurile de gândire recunoscute şi recompensate în profesiile noastre şi prin asta evităm modelele de gândire care ar putea să contravină metodelor şi normelor respectivei profesii. Multe dintre aceste tabuuri sunt meritorii, deoarece asigură ordinea socială şi consolidează progresul. Pe de altă parte însă, ele pot şi să împiedice progresul, prin impunerea unei ortodoxii neproductive. Einstein a vrut să lase în urmă exact această ortodoxie atunci când a încercat să călărească o rază de lumină în experimentele sale mintale.

Regulile culturale îi sunt impuse neocortexului cu ajutorul creierului vechi şi mai ales al amigdalei. Toate gândurile pe care le avem declanşează alte gânduri şi unele dintre ele se referă la pericole asociate. Învăţăm, de exemplu, că încălcarea normelor culturale, chiar şi în

gândurile noastre intime, poate conduce la ostracizare, pe care neocortexul o percepe ca pe o ameninţare la adresa binelui nostru. Dacă nutrim asemenea gânduri, amigdala este stimulată şi generează o stare de frică, iar aceasta, în general, duce la eliminarea gândului.

Pe de altă parte, în vis, aceste tabuuri sunt atenuate, aşa că visăm adeseori chestiuni care sunt interzise cultural, sexual sau profesional. Este ca şi cum creierul nostru şi-ar da seama că, atunci când visăm, nu suntem actori reali în lumea obiectivă. Freud a scris despre acest fenomen, dar a scos în evidenţă şi faptul că ne deghizăm asemenea gânduri periculoase, cel puţin atunci când încercăm să le reapelăm, astfel încât creierul treaz continuă să fie protejat împotriva lor.

Relaxarea tabuurilor profesionale se dovedeşte a fi utilă pentru rezolvarea creativă a problemelor. Eu folosesc în fiecare noapte o tehnică mintală cu ajutorul căreia mă gândesc la o anumită problemă înainte să adorm. Asta îmi declanşează succesiuni de gânduri care vor continua şi în timpul somnului. Când încep să visez, mă pot gândi – *pot visa* – la soluţii pentru problemă eliberat de povara limitărilor profesionale pe care o port pe timpul zilei. Pot apoi accesa aceste vise dimineaţa, când sunt în starea intermediară dintre vis şi veghe, numită uneori „visare lucidă”.5

Este celebră referirea lui Freud la abilitatea de a ajunge la intuiţii cu privire la psihologia persoanei prin interpretarea viselor. Există, desigur, o vastă literatură pe tema tuturor aspectelor acestei teorii, dar ideea fundamentală de obţinere a unor intuiţii privitoare la propria persoană prin examinarea viselor este logică. Visele sunt create de neocortexul nostru şi prin urmare substanţa lor poate să fie revelatoare în privinţa conţinutului şi a conexiunilor care se fac în el. Relaxarea constrângerilor la care este supusă gândirea noastră, constrângeri care există în starea de veghe, este de asemenea utilă pentru dezvăluirea conţinutului

neocortical, pe care altminteri am fi incapabili să-l accesăm direct. Este de asemenea rezonabil să concluzionăm că modelele care ajung în visele noastre reprezintă chestiuni importante pentru noi şi prin urmare sunt indicii privitoare la înţelegerea temerilor şi dorinţelor noastre neîndeplinite.

Rădăcinile unui model

Aşa cum am menţionat mai înainte, în anii 1980 şi 1990 am condus o echipă care a perfecţionat tehnica Modelelor Markov Ascunse Ierarhic pentru recunoaşterea vorbirii umane şi înţelegerea afirmaţiilor în limbaj natural. Aceste cercetări au premers sistemele comerciale răspândite de astăzi care recunosc şi înţeleg ce încercăm să le spunem (sisteme de navigaţie pentru automobile, cu care poţi să vorbeşti, Siri pe iPhone, Google Voice Search şi multe altele). Tehnica pe care am pus-o la punct avea, în esenţă, toate atributele pe care le descriu în Teoria minţii bazate pe recunoaşterea modelelor. Cuprindea o ierarhie de modele în care fiecare nivel superior era mal abstract conceptual decât cel de dedesubtul lui. De exemplu, în recunoaşterea vorbirii, nivelurile Includeau modele de bază ale frecvenţei sunetelor la nivelul cel mai de jos, apoi foneme, apoi cuvinte şi expresii (care erau adeseori recunoscute ca şi cum ar fi fost cuvinte). O parte dintre sistemele noastre de recunoaştere a vorbirii puteau să înţeleagă semnificaţia comenzilor în limbaj natural, iar nivelurile mai înalte includeau structuri de felul expresiilor substantivale şi verbale. Fiecare modul de recunoaştere de model putea recunoaşte o succesiune liniară de modele de la un nivel conceptual inferior. Fiecare imput avea parametri de importanţă, mărime şi variabilitate a mărimii. Existau semnale „descendente" care indicau că era de aşteptat un model de nivel inferior. Voi discuta aceste cercetări mai detaliat în Capitolul 7.

În anii '2003 şi 2004, Jeff Hawkins şi Dileep George, inventatorii lui Palm Pilot, au creat un model cortical ierarhic denumit memorie temporală ierarhică. împreună

cu Sandra Blakesee, specializată pe subiecte ştiinţifice, Hawkins a descris acest model într-un mod elocvent în cartea lor *On Intelligence – Despre inteligenţă*. Hawkins demonstrează foarte convingător uniformitatea algoritmului cortical şi organizarea lui ierarhică bazată pe liste. Între modelul prezentat în On *Intelligence* şi ceea ce prezint eu în această carte există unele deosebiri importante. Aşa cum sugerează şi numele, Hawkins pune accentul pe natura temporală (bazată pe timp) a listelor constituente. Cu alte cuvinte, direcţia listelor este întotdeauna înainte în timp. Explicaţia lui privitoare la modul în care caracteristicile unui model, cum ar fi litera „A" tipărită, are o direcţie în timp se bazează pe mişcarea ochilor. El explică faptul că noi realizăm vizualizări ale imaginilor folosind sacade – mişcări foarte rapide ale ochilor, de care nu suntem conştienţi. Informaţia ce ajunge la neocortex este, prin urmare, nu un set bidimensional de caracteristici, ci, dimpotrivă, o listă ordonată cronologic. Deşi e adevărat că ochii noştri efectuează mişcări foarte rapide, succesiunea în care percep ei caracteristicile unui model cum ar fi litera „A" nu apare întotdeauna într-o ordine temporală sistematică. (De exemplu, sacadele ochilor nu vor înregistra întotdeauna vârful lui „A" înaintea deschiderii de la baza lui.) Mai mult, putem recunoaşte un model vizual care este prezentat cale de doar câteva zeci de milisecunde fără ca asta să însemne prea puţin timp pentru ca sacadele ochilor să-l poată scana. Este adevărat că recognizatorii de model din neocortex stochează modelul ca listă şi că lista este ordonată, dar ordinea nu exprimă în mod necesar timpul. Fireşte, adeseori aşa stau lucrurile, dar este posibil să fie vorba şi despre o ordonare spaţială sau de nivel conceptual superior, aşa cum am discutat anterior.

Cea mai importantă diferenţă o reprezintă setul de parametri pe care i-am inclus pentru fiecare imput în modulul de recunoaştere a modelului – şi în special mărimea şi parametrii de variabilitate a mărimii. În anii

1980, am încercat propriu-zis să recunoaştem vorbirea umană fără acest tip de informaţie. Abordarea respectivă a fost motivată de faptul că lingviştii ne-au spus că durata informaţiei nu era deosebit de importantă. Această perspectivă este ilustrată de dicţionarele care prezintă pronunţia fiecărui cuvânt sub forma unui şir de foneme – de exemplu cuvântul „steep" sub forma [s] [t] [El [p], fără vreo indicaţie referitoare la cât de lung e de aşteptat să fie fiecare fonem. Prin urmare, dacă creăm programe pentru recunoaşterea fonemelor şi apoi întâlnim această succesiune particulară de patru foneme (într-o formă vorbită), ar trebui să putem recunoaşte cuvântul rostit. Sistemul pe care îl construim folosind această abordare a funcţionat într-o anumită măsură, dar nu suficient de bine pentru a putea face faţă unor atribute precum un vocabular bogat, vorbitori multipli şi cuvinte rostite continuu, fără pauze. Când am folosit tehnica Modelelor Mârkov Ascunse Ierarhic pentru a încorpora distribuţia magnitudinilor fiecărui imput, performanţa a crescut imens.

Capitolul 4

NEOCORTEXUL BIOLOGIC

Pentru că lucrurile importante stau în cutii, ai craniul pentru creier, un etii din plastic pentru pieptene şi portofelul pentru bani.

— George Costanza, în episodul „Vizorul invers".

din *Seinfeld*

Acum, pentru prima oară, observăm creierul în funcţiune într-o manieră globală, cu o asemenea claritate, încât ar trebui să fim în stare să descoperim programele complète din spatele acestor magnifice puteri.

— J.G. Taylor, B. Horwitz şi K.J. Friston

Mintea, pe sriurt, lucrează cu datele pe care le primeşte, foarte asemănător cu sculptorul care lucrează cu blocul lui de piatră. Într-un sens, statuia se afla în el de când lumea. Dar mai existau o mie de altele împreună cu ea şi numai sculptorului trebuie să i se mulţumească

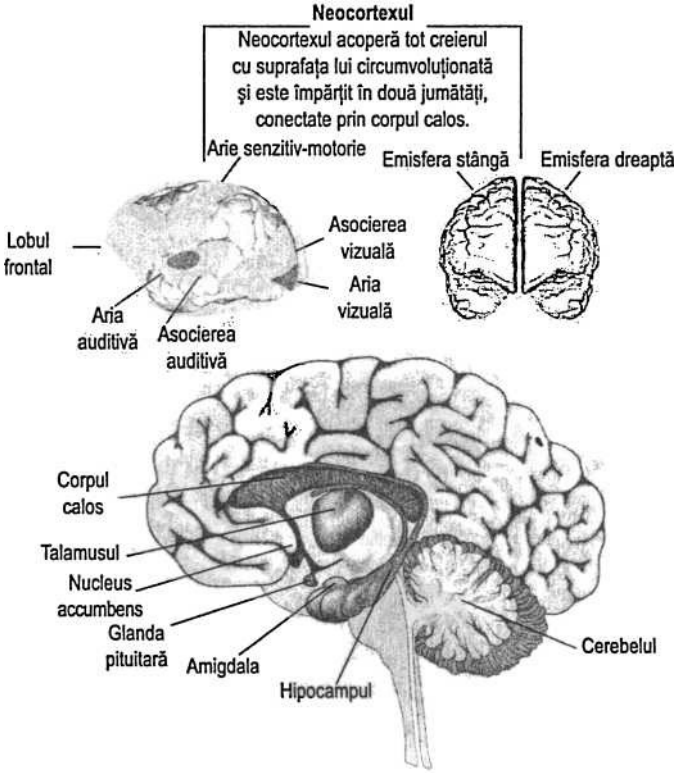
pentru faptul că a degajat-o pe aceasta dintre toate celelalte. Tot astfel, lumile fiecăruia dintre noi, oricât de diferite ar fi punctele noastre de vedere despre ele, sunt înglobate în haosul primordial al senzaţiilor, care a dat *materia* gândului fiecăruia dintre noi atât de fără deosebire. Putem, dacă vrem, prin raţionamente, să descompunem lucrurile până la acea stare de întuneric şi continuitate fără cusături a spaţiului şi a norilor în mişcare de atomi aglomeraţi pe care ştiinţa o numeşte singura lume reală. Cu toate acestea, lumea pe care *noi* o simţim şi în care trăim va fi aceea pe care străbunii noştri şi noi, prin loviturile încet cumulative ale alegerilor, am eliberat-o, asemenea sculptorilor, aruncând pur şi simplu unele părţi din materia ce ne-a fost dată. Alţi sculptori, alte statui, din aceeaşi piatră! Alte minţi, alte lumi, din acelaşi monoton şi inexpresiv haos! Lumea mea nu este decât una dintr-un milion de altele întipărite, la fel de reale pentru cei care ar putea să le abstractizeze. Cât de diferite trebuie să fie lumile în conştiinţa unei furnici, a peştelui-cuţit sau a crabului!

— William James

Este inteligenţa scopul sau măcar unul dintre scopurile evoluţiei biologice? Steven Pinker scrie: „Suntem şovini în privinţa creierului nostru, gândind că el este scopul evoluţiei” 1 şi continuă susţinând că „Asta nu are niciun sens… Selecţia naturală nu face nici pe departe ceva care să indice că se străduieşte pentru inteligenţă. Procesul e propulsat de deosebirile ratelor de supravieţuire şi reproducere ale organismelor care se înmulţesc într-un anumit mediu. Cu timpul, organismele capătă particularităţi care le adaptează pentru supravieţuire şi reproducere în acel mediu. Punct. Nimic nu le trage în vreo altă direcţie decât succesul acolo şi atunci”. Pinker concluzionează că „viaţa e o tufă bogat ramificată, nu o balanţă sau o scară, iar organismele vii se află în vârfurile ramurilor, nu pe treptele inferioare”.

În privinţa creierului uman, Pinker se întreabă dacă

„beneficiile depăşesc costurile”. Dintre costuri, el îl citează pe acela că „creierul [este]

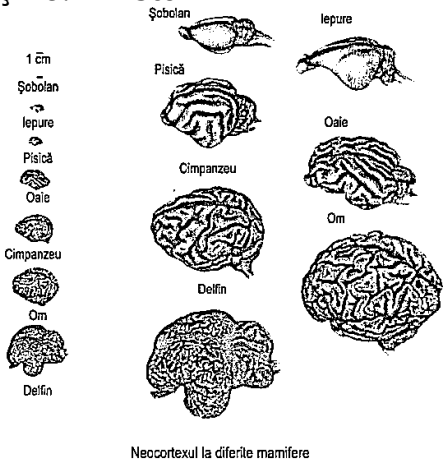


Dispunerea fizică a regiunilor-cheie ale creierului

voluminos. În bazinul femeii, capul supradimensionat al copilului abia dacă poate să aibă loc. Acest compromis de proiectare ucide multe mame la naştere şi impune un mers pivotant care le face pe femei să aibă o mişcare bipedă mai puţin eficientă din punct de vedere biomecanic decât bărbaţii. De asemenea, un cap greu care se susţine pe gât ne face mai vulnerabili la accidente fatale, cum ar fi căzăturile”. Pinker continuă cu enumerarea altor neajunsuri, inclusiv consumul de energie al creierului, timpul mare de reacţie şi procesul îndelungat de învăţare.

Deşi toate aceste afirmaţii sunt exacte la prima vedere (şi cu toate că multe dintre prietenele mele merg pe jos mai bine decât mine), lui Pinker îi scapă, aici, aspectul de ansamblu. Este foarte adevărat că, din punct

de vedere biologic, evoluţia nu are o direcţie specifică. Ea reprezintă o metodă de căutare care încarcă temeinic „ramurile bogat ramificate” ale naturii. La fel de adevărat este şi că schimbările evolutive nu se îndreaptă *neapărat* în direcţia unei inteligenţe sporite – se îndreaptă în *toate* direcţiile. Există



multe exemple de fiinţe pline de succes care au rămas complet neschimbate timp de milioane de ani. (Aligatorii, de exemplu, sunt neschimbaţi de 200 de milioane de ani şi numeroase microorganisme – de mult mai mult timp.) Dar, în cursul încărcării temeinice a nenumăratelor ramuri evolutive, una dintre direcţiile în care se înaintează este o inteligenţă mai mare. Acesta e aspectul relevant pentru scopul acestei discuţii.

Să presupunem că, într-un borcan, avem un gaz albastru. Când scoatem capacul, nu există niciun mesaj care să fie transmis tuturor moleculelor de gaz şi care să le spună: „Hei, băieţi, borcanul nu mai are capacul pus; hai să ne ducem spre gura lui şi să ieşim”. Moleculele pur şi simplu fac ceea ce fac dintotdeauna, adică se mişcă în toate părţile, aparent fără o direcţie anume. Dar, în timp ce fac asta, unele dintre ele, aflate aproape de gura borcanului, chiar ies din el şi, cu timpul, majoritatea moleculelor le urmează exemplul. Din momentul în care a nimerit un mecanism neuronal capabil de învăţare

ierarhizată, evoluţia biologică a constatat că acesta e imens de util pentru unicul obiectiv al evoluţiei, care este supravieţuirea. Beneficiul posesiei unui neocortex a devenit acut în momentul când circumstanţele în rapidă schimbare au necesitat învăţarea rapidă. Specii de toate felurile – plante şi animale – pot învăţa să se adapteze la circumstanţe în schimbare, dar, în lipsa neocortexului, trebuie să folosească procesul de evoluţie genetică. Poate fi nevoie de foarte multe generaţii – mii de ani – pentru ca o specie lipsită de neocortex să înveţe noi comportamente semnificative (sau, în cazul plantelor, noi strategii de adaptare). Avantajul izbitor pentru supravieţuire al neocortexului a constat în faptul că acesta putea să înveţe în decurs de câteva zile. Dacă o specie se confruntă cu o schimbare dramatică a condiţiilor de viaţă şi un membru al speciei inventează, descoperă sau pur şi simplu nimereşte peste un mijloc de adaptare la această schimbare (aceste trei metode fiind, toate, variaţii ale inovaţiei), alţi indivizi observă, învaţă şi copiază metoda, iar aceasta se răspândeşte rapid, viral, la nivelul întregii populaţii. Evenimentul extincţiei cataclismice din Cretacic-Paleogen, din urmă cu aproximativ 65 de milioane de ani, a condus la dispariţia rapidă a multor specii care nu posedau neocortex şi deci nu s-au putut adapta suficient de rapid la mediul brusc modificat. Acesta a fost punctul de turnură pentru mamiferele capabile de neocortex în acţiunea de a-şi asigura nişa ecologică. În acest fel, evoluţia biologică a constatat că proprietatea de învăţare ierarhizată a neocortexului era atât de valoroasă, încât această regiune a creierului a continuat să îşi extindă dimensiunile până când practic a învăluit creierul lui *Homo sapiens*.

Descoperirile din domeniul neuroştiinţelor au stabilit într-un mod convingător rolul-cheie jucat de capacităţile ierarhice ale neocortexului şi în acelaşi timp au adus dovezi în sprijinul Teoriei minţii bazate pe recunoaşterea modelelor. Dovezile constau în numeroase observaţii şi analize, dintre care voi trece o parte în revistă aici.

Psihologul canadian Donald O. Hebb (1904 – 1985) a făcut o primă încercare de a explica baza neurologică a învăţării. În anul 1949, el a descris un mecanism graţie căruia neuronii se modifică fiziologic pe baza experienţei proprii, asigurând astfel temelia pentru învăţare şi pentru plasticitatea creierului: „Să presupunem că persistenţa repetiţiei unei activităţi reverberative (sau «urma») tinde să inducă modificări celulare durabile care îi sporesc stabilitatea… Când axonul unei celule se află suficient de aproape încât să excite o celulă 6 şi când, în mod repetat sau persistent; *celulă A* participă la excitarea celulei 6, are loc un proces de creştere ori o modificare metabolică în una sau în ambele celule, astfel încât eficienţa lui *A*, care este una dintre celulele ce excită celula 6, sporeşte”.2 Această teorie a fost enunţată sub forma „Celulele care excită împreună sunt conectate împreună” şi a devenit cunoscută sub numele de învăţare hebbiană. Aspecte ale teoriei lui Hebb au fost confirmate: este clar că ansamblurile creierului pot să creeze noi conexiuni şi să le întărească pe baza propriei lor activităţi. Practic, putem vedea că neuronii îşi dezvoltă asemenea conexiuni în scanările cerebrale. „Reţelele neuronale” artificiale se bazează pe modelul învăţării neuronale al lui Hebb.

Ipoteza centrală a teoriei lui Hebb este că unitatea fundamentală de învăţare la nivelul, neocortexului o reprezintă neuronul. Teoria minţii bazate pe recunoaşterea modelelor pe care o enunţ în această carte se bazează pe o unitate fundamentală diferită: nu neuronul însuşi, ci mai curând un ansamblu de neuroni, pe care îl estimez la circa o sută. Conectarea şi tăria sinapselor la nivelul fiecărei unităţi sunt relativ stabile şi determinate genetic – respectiv, organizarea la nivelul fiecărui modul de recunoaştere a modelelor este determinată de proiectul genetic. Învăţarea se realizează prin crearea de conexiuni între aceste unităţi şi nu în interiorul lor, dar şi, probabil, prin tăriile sinaptice ale conexiunilor din cadrul unităţilor.

În sprijinul ipotezei că modulul de bază pentru

învăţare este unul compus din zeci de neuroni vine specialistul elveţian în neuroştiinţă Henry Markram (născut în 1962), căruia îi descriu ambiţiosul Proiect Blue Brain de simulare a întregului creier uman în Capitolul 7. Într-o lucrare din 2011, el dezvăluie faptul că, atunci când a scanat şi analizat neuroni de mamifere, potrivit propriei formulări „căutam dovezi ale ansamblurilor hebbiene la nivelul cel mai elementar al cortexului”: Ce a descoperit în schimb, scrie el, au fost „ansambluri evazive [ale căror] conectivităţi şi greutăţi sinaptice sunt extrem de previzibile şi restricţionate”. Markram concluzionează că „aceste descoperiri implică faptul că experienţa nu poate modela cu uşurinţă conexiunile sinaptice ale ansamblurilor” şi speculează că „pentru percepţie, [ele] au nevoie de blocuri de cunoaştere înnăscute, de genul pieselor de Lego, şi că achiziţia de amintiri implică 6 combinare a acestor blocuri în construcţii complexe”. Markram continuă:

De decenii întregi a fost comunicată existenţa ansamblurilor neuronale funcţionale, dar dovezi directe ale existenţei aglomerărilor de neuroni conectaţi sinaptic… au lipsit… De vreme ce aceste ansambluri sunt toate similare ca topologie şi greutate sinaptică, nefiind modelate de vreo experienţă specifică, noi considerăm că sunt ansambluri înnăscute… Experienţa joacă un rol minor în determinarea conexiunilor sinaptice şi a greutăţii lor în cadrul acestor ansambluri… Studiul nostru a condus la descoperirea de dovezi [ale existenţei] ansamblurilor înnăscute de genul pieselor Lego, compuse din câteva zeci de neuroni… Conexiunile dintre ansambluri pot să le combine pe acestea în superansambluri la nivelul unui strat neocortical, apoi în ansambluri de ordin superior într-o coloană corticală, ba chiar în superansambluri într-o regiune a creierului şi, în cele din urmă, într-un ansamblu de ordinul cel mai înalt, reprezentat de întregul creier… Achiziţia de amintiri este foarte asemănătoare construirii cu piese Lego. Fiecare ansamblu e echivalent cu o piesă

Lego care conţine cunoştinţe elementare înnăscute referitoare la modul de procesare, de percepţie şi de răspuns la lume… Când se reunesc, mai multe blocuri diferite pot, prin urmare, să formeze o combinaţie unică a acestor percepţii înnăscute, care reprezintă cunoştinţele şi experienţele specifice ale individului.3

„Piesele Lego” pe care le propune Markram sunt pe deplin compatibile cu modulele de recunoaştere a modelelor pe care le-am descris eu. Într-o comunicare prin e-mail, Markram a descris aceste „piese Lego” ca fiind „conţinut împărtăşit şi cunoştinţe înnăscute”.4 Eu aş afirma că scopul acestor module este să recunoască modele, să le memoreze şi să le anticipeze pe baza unor modele parţiale. A se remarca faptul că estimarea lui

Markram potrivit căreia fiecare modul ar conţine „câteva zeci de neuroni” se bazează numai pe stratul V al neocortexului. Stratul V este într-adevăr bogat în neuroni, dar, raportat la numărul obişnuit de neuroni estimaţi pe stratul VI, asta s-ar traduce printr-un ordin de mărime de aproximativ 100 de neuroni per modul, cifră aflată în concordanţă cu estimările mele.

Interconexiunile consistente şi aparentul caracter modular al neocortexului au fost observate de ani de zile, dar studiul descris anterior e primul care demonstrează stabilitatea acestor module atunci când creierul îşi derulează procesele dinamice.

Un alt studiu recent, realizat de data asta la General Hospital din Massachusetts, finanţat de National Institutes of Health şi de National Science Foundation şi dat publicităţii într-o ediţie din luna martie a anului 2012 a revistei *Science*, vorbeşte de asemenea despre o structură de conexiuni la nivelul întregului neocortex.5 Articolul arată că, la nivelul neocortexului, conexiunile respectă, un model de grilă similar celui al străzilor ordonate dintr-un oraş: „În esenţă, structura de ansamblu a creierului ajunge să se asemene cu Manhattan-ul, în care ai o hartă 2-D a străzilor şi o axă, un lift, care intră în a treia dimensiune”,

a scris Van J. Wedeen, specialist în neuroştiinţă şi fizician la Harvard, şeful studiului.

Într-un podcast al revistei *Science*, Wedeen a descris semnificaţiile cercetărilor: „A fost un studiu asupra structurii tridimensionale a căilor situate la nivelul creierului. Când oamenii de ştiinţă s-au gândit la căile din creier în ultima sută de ani, imaginea tipică sau modelul care le-a venit în minte a fost acela al unor căi asemănătoare cu spaghetele dintr-un bol – trasee separate, cu modele spaţiale prea puţin legate unele de altele. Folosind rezonanţa magnetică, am putut investiga experimental această chestiune. Şi am descoperit că, în loc să existe căi aranjate întâmplător sau independente, toate traseele din creier corespund unei structuri extraordinar de simple. În esenţă, ansamblul seamănă cu un cub. Căile se întind pe cele trei direcţii perpendiculare şi, pe fiecare dintre aceste trei direcţii, sunt perfect paralele între ele şi sunt aranjate în serii. Deci, în locul unui bol de spaghete independente, observăm că, că la nivelul creierului, conectivitatea este, într-un sens, o unică structură coerentă”.

În vreme ce studiul lui Markram prezintă un modul de neuroni care se repetă la nivelul întregului neocortex, studiul lui Wedeen demonstrează existenţa unui model remarcabil de ordonat de conexiuni între module. Creierul porneşte de la un număr foarte mare de „conexiuni în aşteptare”, la care se pot lega modulele de recunoaştere a modelelor. Prin urmare, dacă un anumit modul vrea să se conecteze cu altui, nu trebuie să aibă un axon din unul şi o dendrită din celălalt care să acopere distanţa fizică dintre ele. Poate, pur şi simplu, să acceseze una dintre aceste conexiuni axonice în aşteptare şi să se lege automat la capetele fibrei. Aşa cum scriu şi Wedeen şi colaboratorii săi: „Căile creierului respectă un plan de bază stabilit de… embriogeneza timpurie. Prin urmare, căile creierului ajuns la maturitate reflectă aceste trei gradiente primordiale deformate fizic de dezvoltare”. Cu alte cuvinte, pe măsură

ce învăţăm şi avem experienţe, modulele de recunoaştere a modelelor din neocortex se conectează la aceste legături preexistente care au fost create pe când eram embrioni.

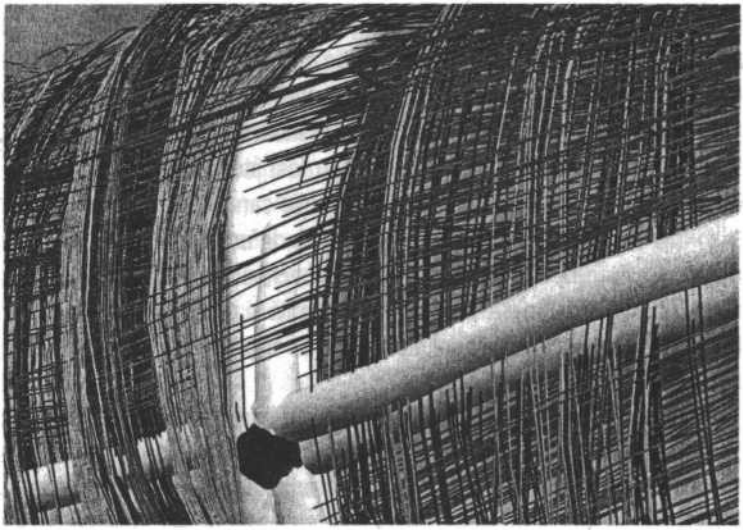
Există un cip electronic, numit *Field Programmable Gate Array* (*FPGA*) – serie de porţi de câmp programabile –, care se bazează pe un principiu asemănător. Cipul conţine milioane de module care implementează funcţii logice şi creează conexiuni în aşteptare. În momentul utilizării, conexiunile respective nu sunt nici activate, nici dezactivate (prin semnale electronice) în vederea implementării unei anumite capacităţi.

La nivelul neocortexului, aceste conexiuni la mare distanţă care rămân nefolosite sunt eventual tăiate. Este unul dintre motivele pentru care adaptarea unei regiuni învecinate a neocortexului în vederea compensării uneia lezate nu este la fel de eficientă ca regiunea originală. Potrivit studiului lui Wedeen, conexiunile iniţiale sunt extrem de ordonate şi repetitive, exact la fel ca modulele propriu-zise, iar modelul lor în grilă este folosit pentru „ghidarea conectivităţii” la nivelul neocortexului. Acest model a fost identificat la toate creierele de primate şi de oameni care au fost studiate şi s-a dovedit a fi evident la nivelul neocortexului de la regiunile care se ocupă de modelele senzoriale timpurii, până la cele responsabile de emoţiile de nivel superior. Articolul lui Wedeen din revista *Science* concluzionează că „structura de grilă a căilor cerebrale este universală, coerentă şi continuă pe cele trei axe principale de dezvoltare”. Acest fapt indică un algoritm comun pentru toate funcţiile neocortexului.

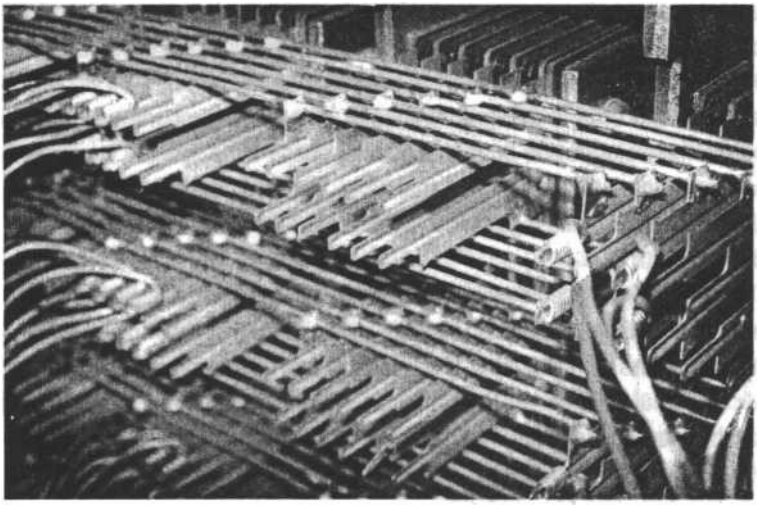
Este cunoscut de mult faptul că măcar anumite regiuni ale neocortexului sunt ierarhice. Regiunea cel mai bine studiată este cortexul vizual, separat în arii denumite V1, V_2 şi MT (cunoscută de asemenea şi ca V_5). Pe măsură ce înaintăm în arii superioare din această regiune („superioare” în sensul procesării conceptuale, nu fizic, deoarece neocortexul are mereu o grosime egală cu un singur recognizator de modele), proprietăţile care pot fi

recunoscute

Structura extraordinar de regulată a conexiunilor iniţiale de la nivelul neocortexului, descoperită în cadrul unui studiu al National Institutes of Health



O altă vedere a structurii regulate de tip grilă a conexiunilor de la nivelul neocortexului



Structura de tip grilă descoperită la nivelul neocortexului este remarcabil de asemănătoare cu ceea ce se numeşte comutator încrucişat, care e folosiţ în circuitele integrate şi în plăcile de circuite devin tot mai abstracte. Aria V1 recunoaşte muchiile elementare şi formele primitive. Aria V_2 recunoaşte contururile, diferenţele dintre imaginile prezentate de fiecare ochi, orientarea spaţială şi posibilitatea ca o porţiune din imagine să facă parte din obiect sau din fundal ori ba.6 Regiunile de nivel superior ale neocortexului recunosc concepte precum identitatea obiectelor şi a chipurilor şi mişcările lor. De asemenea, este de multă vreme cunoscut faptul că orice comunicare prin intermediul acestei ierarhii se face atât în sens ascendent, cât şi descendent şi că semnalele pot fi atât excitatoare, cât şi inhibitoare. Tomaso Poggio (născut în 1947), specialist în neuroştiinţă la MIT, a studiat în detaliu vederea la nivelul creierului uman, iar cercetările sale din ultimii treizeci şi cinci de ani au fost esenţiale pentru localizarea învăţării ierarhice şi pentru recunoaşterea modelelor la nivelurile „timpurii” (cele mai reduse conceptual) ale neocortexului vizual.7

Ceea ce înţelegem noi cu privire la nivelurile ierarhice inferioare ale neocortexului vizual Corespunde Teoriei minţii bazate pe recunoaşterea modelelor pe care

am descris-o în capitolul anterior, iar observaţiile privitoare la natura ierarhică a procesării neocorticale s-au extins recent cu mult dincolo de aceste niveluri. Daniel J. Felleman, profesor de neurobiologie la Universitatea Texas, şi colegii săi au identificat „organizarea ierarhică a cortexului cerebral… [în 25 de arii ale neocortexului", care includ atât ariile vizuale, cât şi pe cele de nivel superior care combină modelele primite de la mai multe organe de simţ. Avansând în ierarhia neocortexului, procesarea modelelor devine tot mai abstractă, ocupă zone spaţiale tot mai largi şi presupune intervale de timp mai lungi. Pentru fiecare conexiune, ei au descoperit în ierarhie legături de comunicare atât în sens ascendent, cât şi descendent.8

Cercetările recente ne permit să ne lărgim substanţial aceste observaţii înspre regiuni care depăşesc cu mult cortexul vizual şi chiar înspre ariile de asociere, care combină inputul primit de la mai multe simţuri. Un studiu publicat în anul 2008 de Un Hasson, profesor de psihologie la Princeton, şi de colaboratorii săi demonstrează că fenomenele observate la nivelul cortexului vizual se produc în numeroase arii ale neocortexului: „Este bine stabilit faptul că neuronii de pe căile corticale vizuale au câmpuri receptoare spaţiale din ce în ce mai mari. Este un principiu de organizare elementar al sistemului vizual… Evenimentele din lumea reală apar nu doar în regiuni ample ale spaţiului, ci şi în perioade extinse de timp. Prin urmare, am formulat ipoteza că ar trebui să existe o ierarhie analoagă celei descoperite pentru mărimea câmpurilor receptoare spaţiale şi pentru caracteristicile de răspuns temporal ale diferitelor regiuni ale creierului". Şi exact asta au descoperit, ceea ce le-a permis să concluzioneze că, „similar ierarhiei corticale cunoscute a câmpurilor receptoare spaţiale, în creierul omului există o ierarhie a ferestrelor receptoare temporale progresiv crescătoare".9

Cel mai convingător argument privitor la universalitatea procesării la nivelul neocortexului îl

reprezintă dovada omniprezentă a plasticităţii (nu doar învăţare, ci şi interschimbabilitate); cu alte cuvinte, o regiune e capabilă să facă munca altor regiuni, fapt ce implică existenţa unui algoritm comun la nivelul întregului neocortex. O parte foarte importantă a cercetărilor din domeniul neuroştiinţei s-a concentrat asupra stabilirii regiunilor din neocortex care au legătură cu anumite tipuri de modele.

Tehnica clasică de determinare a acestor legături a luat în considerare creierele care au suferit leziuni în urma unor traumatisme sau atacuri cerebrale şi asocierea funcţiilor pierdute cu regiunile specifice lezate. De exemplu, atunci când observăm că o persoană care a suferit o leziune de dată recentă în regiunea girusului fusiform are dificultăţi în recunoaşterea chipurilor, dar îşi poate identifica semenii după voce şi modele de limbaj, putem emite ipoteza că respectiva regiune are legătură cu recunoaşterea chipurilor. Ipoteza fundamentală a fost aceea că fiecare dintre regiunile cu pricina este concepută să recunoască şi să proceseze un anumit tip de model. Anumite regiuni fizice au fost asociate cu anumite tipuri de modele, deoarece, în condiţii normale, aşa circulă informaţia. Dar, când fluxul normal de informaţie este tulburat dintr-un motiv oarecare, o altă regiune a neocortexului se vădeşte a fi capabilă să intervină şi să preia funcţia.

Plasticitatea a fost larg observată de neurologi, care au sesizat că pacienţii cu leziuni cerebrale survenite în urma unui traumatism sau a unui atac cerebral pot reînvăţa aceleaşi abilităţi într-o altă regiune a neocortexului. Probabil, exemplul cel mai dramatic de plasticitate este de găsit într-un studiu din anul 2011 al specialistului american în neuroştiinţă Marina Bedny şi al colaboratorilor săi. Studiul vizează situaţia cortexului vizual la persoanele oarbe congenital. Opinia generală era că straturile timpurii ale cortexului vizual, cum ar fi V1 şi V_2, se ocupă în mod inerent de modelele inferioare (ca de

pildă muchiile şi curbele), în vreme ce cortexul frontal (acea nouă regiune evolutivă a cortexului pe care o avem în spatele frunţilor noastre unic de largi) se ocupă în mod inerent de modelele mai complexe şi mai subtile ale limbajului şi de conceptele abstracte. Dar, aşa cum au descoperit Bedny şi colaboratorii săi, „se crede că omul posedă la nivelul lobilor frontali şi temporali stâng şi drept regiuni cerebrale evoluate, capabile în mod unic de procesarea limbajului. Pe de altă parte însă, persoanele oarbe congenital activează şi cortexul vizual pentru anumite funcţii verbale. Noi aducem dovezi că această activitate a cortexului vizual reflectă, de fapt, procesarea limbajului. Am descoperit că, la persoanele oarbe congenital, cortexul vizual stâng se comportă similar regiunilor clasice ale limbajului… Concluzionăm că regiunile creierului care se consideră că au evoluat în scopul vederii pot adopta procesarea limbajului ca rezultat al unei experienţe timpurii".10 Gândeşte-te numai, cititorule, la implicaţiile acestui studiu: înseamnă că regiuni ale neocortexului care sunt relativ depărtate sub raport fizic şi care au fost considerate şi foarte diferite din punct de vedere conceptual (sugestii vizuale primitive *versus* concepte abstracte de limbaj) folosesc în esenţă acelaşi algoritm. Regiunile care procesează aceste tipuri disparate de modele se pot substitui reciproc.

Daniel E. Feldman, specialist în neuroştiinţă la Universitatea California, Berkeley, a scris o cuprinzătoare recenzie pe tema a ceea ce el a numit „mecanismele sinaptice pentru plasticitatea neocortexului" şi a descoperit dovezi referitoare la acest tip de plasticitate la nivelul întregului neocortex. El notează că „plasticitatea îi permite creierului să înveţe şi să-şi reamintească modele din lumea senzorială, să rafineze mişcări… şi să-şi recapete funcţia după leziuni". El adaugă şi că această plasticitate este făcută posibilă de „modificări structurale, inclusiv compunerea, îndepărtarea şi remodelârea morfologică a sinapselor corticale şi a crestelor

dendritice”.11

Un alt exemplu uluitor de plasticitate neocorticală (şi, prin urmare, venind în sprijinul uniformităţii algoritmului neocortical) a fost demonstrat recent de oamenii de ştiinţă de la Universitatea California, Berkeley. Ei au montat serii de microelectrozi pentru culegerea de semnale cerebrale dintr-o regiune a cortexului motor al şoarecilor care controlează mişcările mustăţilor. Cercetătorii au conceput experimentul astfel încât şoarecii să obţină o recompensă dacă îşi controlau aceşti neuroni pentru a genera un anumit model mintal fără să-şi mişte mustăţile. Modeiul necesar pentru obţinerea recompensei presupunea o sarcină mintală pe care neuronii frontali ai şoarecilor nu ar fi îndeplinit-o în mod normal. Însă şoarecii s-au dovedit capabili să îndeplinească această sarcină mintală gândind în esenţă cu neuronii motori pe care îi decuplau mintal de la controlul mişcărilor motorii.12 Concluzia este că neocortexul motor, regiune a neocortexului responsabilă de coordonarea mişcărilor musculare, foloseşte şi el un algoritm neocortical standard.

Pe de altă parte, există mai multe motive pentru care o abilitate sau un domeniu de cunoaştere reînvăţat cu ajutorul unei noi arii a neocortexului, care o înlocuia pe una lezată, nu este la fel de bună ca şi cea originală. În primul rând, deoarece a fost nevoie de o viaţă întreagă pentru învăţarea şi perfecţionarea unei anumite abilităţi, reînvăţarea într-o altă arie a neocortexului nu poate genera imediat aceleaşi rezultate. Mai important, acea altă arie a neocortexului nu a stat degeaba, pe post de rezervă pentru regiunea lezată. Şi ea a îndeplinit funcţii vitale, aşa că va ezita să renunţe la modelele sale neocorticale pentru a compensa regiunea lezată. Ea poate începe prin a elibera o parte dintre copiile redundante ale modelelor sale, dar, prin aceasta, va degrada în mod subtil abilităţile existente şi nu va elibera atât de mult spaţiu cortical cât folosiseră iniţial abilităţile care sunt reînvăţate acum.

Mai există şi un al treilea motiv pentru care

plasticitatea are limitele ei. De vreme ce la majoritatea oamenilor tipurile de modele se transmit prin regiuni specifice (de exemplu, chipurile sunt procesate de girusul fusiform), aceste regiuni au ajuns să fie optimizate (de evoluţia biologică) pentru respectivele tipuri de modele. După cum voi arăta în Capitolul 7, am descoperit aceleaşi rezultate în creaţiile noastre neocorticale digitale. Am putea recunoaşte vorbirea cu sistemele noastre de recunoaştere a caracterelor şi viceversa, dar sistemele de vorbire au fost optimizate pentru vorbire şi, în mod similar, sistemele de recunoaştere a caracterelor au fost optimizate pentru caracterele tipărite, deci.ar exista, o anumită diminuare a performanţei dacă l-am substitui pe unul cu celălalt. Am folosit practic algoritmi evolutivi (genetici) pentru realizarea acestei optimizări, o simulare a ceea ce biologia face în mod natural. Dat fiind faptul că chipurile au fost transmise prin girusul fusiform la majoritatea oamenilor timp de sute de mii de ani (sau mai mult), evoluţia biologică a avut timpul necesar pentru a dezvolta o capacitate favorabilă procesării unor asemenea modele în acea regiune. Girusul fusiform foloseşte aceiaşi algoritmi de bază, dar este orientat către chipuri. Aşa cum a scris specialistul olandez în neuroştiinţă Randal Koene, „neocortexul este foarte uniform, fiecare coloană sau minicoloană putând să facă, în principiu, tot ceea ce face o alta”.13

Cercetări substanţiale susţin observaţia că modulele de recunoaştere a modelelor se conectează pe baza modelelor la care sunt expuse. De exemplu, specialistul în neuroştiinţă Vi Zuo şi colegii ei au urmărit felul cum noi „spini dendritici” formau conexiuni între celulele nervoase în timp ce şoarecii învăţau o nouă abilitate (să întindă laba printr-o fantă ca să apuce o sămânţă).14 Cercetătorii de la Institutul Salk au descoperit că această autoconectare fundamentală a modulelor neocortexului este controlată, aparent, de numai câteva gene. Aceste gene şi această metodă de autoconectare sunt de asemenea uniforme la

nivelul neocortexului.15

Multe alte studii documentează aceste atribute ale neocortexului, dar să rezumăm ce putem observa din literatura neuroştiinţei şi din propriile experimente mintale. Unitatea de bază a neocortexului este un modul de neuroni al căror număr îl estimez la circa o sută. Aceste module sunt întreţesute în fiecare coloană a neocortexului, astfel încât un modul în parte nu este vizibil distinct. Modelul conexiunilor şi tăriilor sinaptice din cadrul fiecărui modul este relativ stabil. Tocmai conexiunile şi tăriile sinaptice dintre module sunt cele care reprezintă învăţarea.

La nivelul neocortexului există ceva de ordinul unui cvadrilion (1015) de conexiuni, dar numai în jur de 25 de biţi de informaţie creativă la nivelul genomului (după compresia fără pierdere) 16, astfel încât conexiunile propriu-zisé nu pot fi în niciun caz predeterminate genetic. Este posibil ca o parte din această învăţare să fie produsul interogării creierului vechi de către neocortex, dar şi asta ar reprezenta doar un volum relativ mic de Informaţie. Conexiunile dintre module sunt generate în ansamblul lor în urma experienţei (prin cultivare, mai degrabă decât de la natură).

Creierul nu are suficientă flexibilitate, astfel încât fiecare modul de recunoaştere a modelelor se poate conecta pur şi simplu cu orice alt modul (aşa cum ne putem programa cu uşurinţă computerele sau cum ne putem lega la web). Trebuie să fie realizată o conexiune fizică propriu-zisă, formată dintr-un axon legat de o dendrită. Toţi ne începem viaţa cu un vast stoc de posibile conexiuni neuronale. După cum demonstrează studiul lui Wedeen, aceste conexiuni sunt organizate de o manieră hiperrepetitivă şi ordonată. Conexiunea terminală cu aceşti axoni în aşteptare se produce pe baza modelelor pe care le-a recunoscut fiecare recognizator de modele din neocortex. Conexiunile neutilizate sunt în cele din urmă tăiate. Aceste conexiuni sunt construite ierarhic,

reflectând ordinea ierarhică naturală a realităţii. Acesta este punctul-forte cheie al neocortexului.

Algoritmul de bază al modulelor de recunoaştere a modelelor din neocortex este acelaşi pentru întregul neocortex, de la modulele de „nivel redus”, care se ocupă de cele mai elementare modele senzoriale, până la modulele de „nivel superior”, care recunosc cele mai abstracte concepte. Vasta evidenţă a plasticităţii şi interschimbabilităţii regiunilor neocortexului este mărturia acestei importante observaţii. Există o anumită optimizare a regiunilor care se ocupă cu anumite tipuri de modele, dar acest efect este de ordin secund – algoritmul fundamental e universal.

Semnalele se transmit ascendent şi descendent în ierarhia conceptuală. Un semnal cu sens ascendent înseamnă „Am detectat un model”. Un semnal cu sens descendent înseamnă „Mă aştept să apară modelul tău” şi este, în esenţă, o predicţie. Atât semnalele ascendente, cât şi cele descendente pot fi ori excitatoare ori inhibitoare.

Fiecare model e poziţionat el însuşi într-o anumită ordine, deci este greu de inversat. Chiar dacă un model pare să aibă aspecte multidimensionale, el e reprezentat de o succesiune unidimensională de modele de nivel inferior. Modelul este o succesiune ordonată de alte modele, astfel încât fiecare recognizator este în mod inerent recursiv. Pot să existe numeroase nivele ierarhice.

În modelele pe care le învăţăm se află o extraordinară redundanţă, mai ales în cele importante. Recunoaşterea modelelor (cum ar fi obiectele şi chipurile obişnuite) apelează la aceleaşi mecanisme ca amintirile, care sunt simple modele pe care le-am învăţat. Ele sunt de asemenea stocate ca succesiuni de modele – adică sunt în esenţă poveşti. Acest mecanism este folosit şi pentru învăţarea şi realizarea mişcării fizice în lume. Redundanţa modelelor ne permite să recunoaştem obiecte, oameni şi idei chiar şi atunci când aparvariaţiuni, iar contextele sunt diferite. De asemenea, dimensiunea şi variabilitatea

mărimii parametrilor îi permit neocortexului să codifice variaţia de magnitudine comparativ cu diferite unităţi de mărime (durata, în cazul sunetului). Un mod de a codifica aceşti parametri de magnitudine îl constituie modelele multiple cu numere diferite de inputuri repetate. Aşa, de exemplu, pot să existe modele pentru cuvântul rostit „steep” cu numere diferite de vocale lungi repetate [El, fiecare cu parametrul de importanţă stabilit la un nivel moderat, care să indice faptul că repetarea lui [El este variabilă. Această abordare nu e matematic echivalentă cu faptul de a avea parametri de mărime expliciţi şi nu dă rezultate la fel de bune în practică, însă reprezintă o abordare a codificării magnitudinii. Cea mai elocventă dovadă pe care o avem în privinţa acestor parametri o reprezintă faptul că ei sunt necesari sistemelor noastre de inteligenţă artificială pentru obţinerea unor niveluri de acurateţe apropiate de cele umane.

Rezumatul de mai înainte constituie concluziile pe care le-am tras din experimentele despre care am discutat anterior şi din experimentele mintale despre care de asemenea am discutat. Susţin în continuare că modelul pe care l-am prezentat este singurul care poate satisface toate constrângerile impuse de cercetările şi experimentele noastre mintale.

În final, mai există o mărturie care poate fi coroborată; Tehnicile pe care le-am perfecţionat pe parcursul ultimelor decenii în domeniul inteligenţei artificiale pentru recunoaşterea şi procesarea inteligentă a fenomenelor din lumea reală (cum ar fi limbajul uman vorbit şi cel scris) şi pentru înţelegerea documentelor în limbaj natural se dovedesc a fi similare din punct de vedere matematic modelului pe care l-am prezentat mai înainte. Sunt de asemenea exemple ale Teoriei minţii bazate pe recunoaşterea modelelor. Domeniul inteligenţei artificiale nu a încercat în mod explicit să copieze creierul, dar, cu toate acestea, a ajuns la tehnici în esenţă echivalente.

Capitolul 5

CREIERUL VECHI

Am un creier vechi, dar o memorie extraordinară.

— Al Lewis iată-ne în mijlocul acestei noi lumi cu creierul nostru primitiv acordat la viaţa simplă de cavernă, dar cu forţe extraordinare la dispoziţia noastră, pe care suntem destul de deştepţi să le descătuşăm, însă cărora nu le putem înţelege consecinţele.

— Albert Szent-Gyorgyi

Creierul nostru vechi – cei pe care l-am avut înainte de a fi mamifere – nu a dispărut. Continuă, fireşte, să asigure o mare parte din motivaţia noastră pentru căutarea recompenselor şi evitarea pericolelor. Aceste scopuri sunt modulate, în schimb, de neocortexul nostru, care domină creierul omenesc atât ca masă, cât şi ca activitate.

Animalele au trăit şi au supravieţuit fără neocortex şi toate animalele care nu sunt mamifere continuă să o facă în zilele noastre. Putem considera neocortexul omului ca fiind un mare sublimator – astfel, motivaţia noastră primitivă pentru evitarea unui animal mare de pradă poate fi transformată de neocortexul de azi în îndeplinirea unei sarcini pentru impresionarea şefului; marea vânătoare poate să devină scrierea unei cărţi, să spunem despre minte; iar urmărirea reproducerii poate să devină obţinerea recunoaşterii publice sau decorarea apartamentului. (Ei bine, această din urmă motivaţie nu este totdeauna chiar atât de ascunsă.)

Neocortexul este bun, de asemenea, la a ne ajuta să rezolvăm probleme, deoarece poate crea modele precise ale lumii, reflectându-i adevărata natură ierarhică. Însă cel care ne pune în faţa tuturor acestor probleme este creierul vechi. Desigur, ca orice birocraţie inteligentă, neocortexul rezolvă adeseori problemele care-i sunt alocate prin redefinirea lor. Pe această notă, să trecem în revistă procesarea informaţiei la nivelul creierului vechi.

Calea senzorială

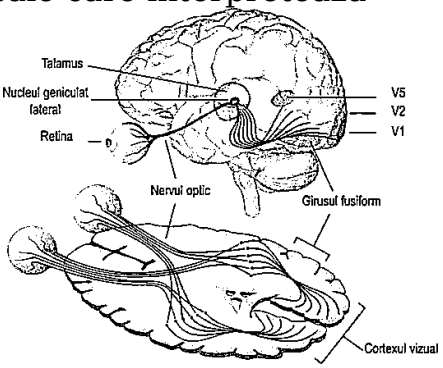
Tablourile, propagate prin mişcare de-a lungul fibrelor nervului optic la creier, sunt cauza vederii.

— Isaac Newton

Fiecare dintre noi trăieşte înăuntrul universului – închisoarea – propriului său creier. Din el proiectează milioane de fibre nervoase senzoriale fragile, în grupuri adaptate unic încercării stărilor energetice ale lumii din jurul nostru: căldura, lumina, forja şi compoziţia chimică. Asta e tot ce ştim despre el în mod direct; în rest, totul este deducţie logică.

— Vernon Mountcastle1

Cu toate că trăim experienţa iluziei că primim imagini de înaltă rezoluţie de la ochi, ceea ce transmite de fapt la creier nervul optic este doar o serie de schiţe şi indicii cu privire la punctele de interes din câmpul nostru vizual. Pe urmă, în esenţă, visăm lumea din amintirile corticale care interpretează

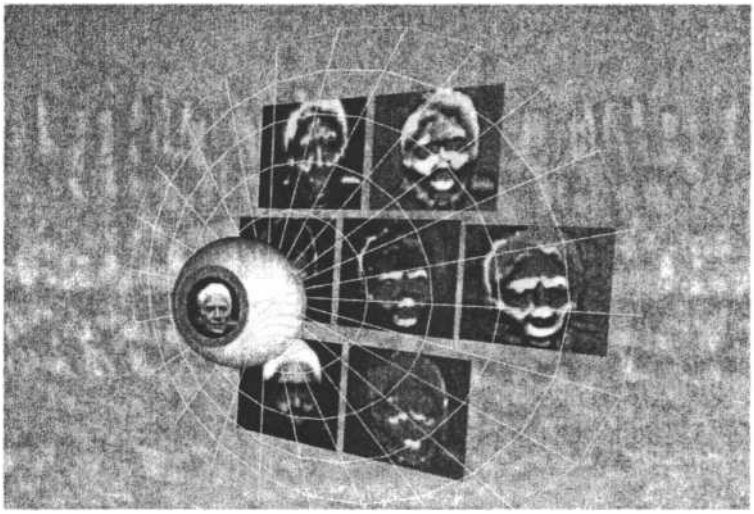


Căile vizuale la nivelul creierului o serie de filme cu rate ale datelor foarte reduse, care sosesc pe canale paralele. Într-un studiu publicat în revista *Nature*, Frank S. Werblin, profesor de biologie moleculară şi celulară la Universitatea California din Berkeley, şi Boton Roska, doctor în medicină, au arătat că nervul optic conţine între zece şi douăsprezece canale de output, fiecare dintre ele transportând doar un mic volum de informaţie despre o scenă dată.2 Un grup din ceea ce se numeşte celule ganglionare trimite informaţii referitoare numai la muchii (modificări ale contrastului). Un alt grup detectează doar

ariile întinse de culoare uniformă, în vreme ce un al treilea grup este sensibil numai la fundăturile din spatele obiectelor de interes.

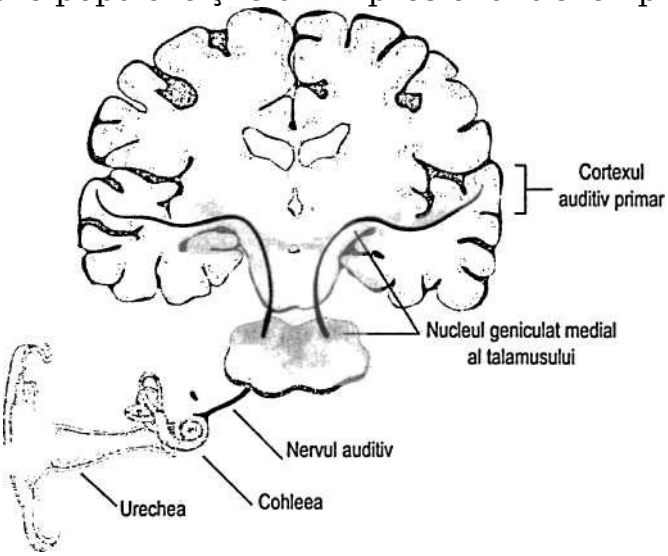
„Chiar dacă credem că vedem lumea cât se poate de complet, ceea ce recepţionăm în realitate sunt numai indicii, muchii în spaţiu şi timp”, spune Werblin. „Aceste 12 imagini ale lumii reprezintă toată informaţia pe care o vom deţine vreodată referitor la ceea ce există, iar din aceste 12 imagini, care sunt atât de sporadice, noi reconstruim bogăţia lumii vizuale. Sunt curios cum a putut natura să aleagă aceste 12 imagini simple şi cum este posibil ca ele să fie suficiente pentru a ne asigura întreaga informaţie de care s-ar părea că avem nevoie”.

Această reducere a datelor reprezintă ceea ce noi, în domeniul inteligenţei artificiale, numim „codificare sporadică”. În crearea sistemelor



Şapte dintre cele douăsprezece „filme” cu conţinut redus de date transmise de nervul optic la creier artificiale, am descoperit că eliminarea unei mari părţi din informaţiile de imput şi reţinerea doar a celor mai sugestive detalii asigură rezultate superioare. Altminteri, abilitatea limitată de procesare a informaţiilor la nivelul neocortexului (biologică sau de altă natură) ajunge să fie depăşită.

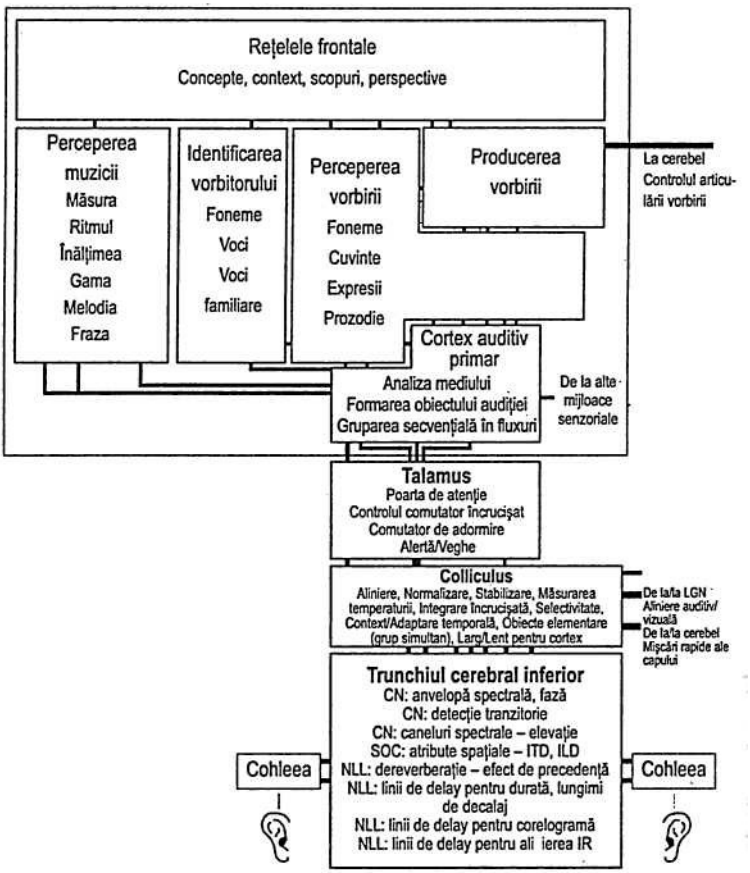
Lloyd Watts, împreună cu echipa lui de cercetare de la Audience, Inc, a creat un model minuţios al procesării informaţiei auditive de către cohleea umană prin intermediul regiunilor subcorticale şi apoi prin etajele inferioare ale neocortexului.3 Această echipă a pus la punct o tehnologie de cercetare care extrage 600 de benzi de frecvenţă diferite (60 la o octavă) din sunet. Această valoare se apropie mai mult de estimarea de 3.000 de benzi extrase de cohleea umană (în comparaţie cu sistemele comerciale de recunoaştere a vorbirii, care folosesc numai 16 până la 32 de benzi). Utilizând două microfoane şi modelul detaliat (plus rezoluţia spectrală înaltă) de procesare auditivă, Audience a creat o tehnologie comercială (cu o rezoluţie spectrală oarecum mai redusă decât sistemul de cercetare) care elimină eficient zgomotul de fundal din conversaţii. Această tehnologie este folosită actualmente de multe telefoane celulare populare şi e un impresionant exemplu



Calea auditivă la nivelul creierului de produs comercial bazat pe înţelegerea modului în care sistemul auditiv al omului este capabil să se concentreze asupra unei surse sonore de interes.

Imput-urile de la corp (estimate la sute de megabiţi

pe secundă), inclusiv de la nervii din piele, muşchi, organe şi din alte arii, ajung la nivelul superior al măduvei spinării. Aceste mesaje implică mai mult decât comunicarea referitoare la atingere; în plus, transportă informaţii referitoare la temperatură.



Un model simplificat al procesării auditive atât la nivelul ariilor subcorticale (arii anterioare neocortexului), cât şi la nivelul neocortexului, creat de Audience Inc. Imagine adaptată după L. Watts, *Reverse-Engineering the Human Auditory Pathway*, în J. Liu şi colab. (sub redacţia), WCC12012 (Berlin: Springer-Veriag, 2012), pag. 49

nivelul acizilor (de exemplu, acidul lactic din muşchi), deplasarea alimentelor prin tractul gastrointestinal şi multe alte semnale. Aceste date sunt procesate prin

trunchiul cerebral şi creierul mijlociu. Celule-cheie numite neuroni lamina 1 creează o hartă a corpului, reprezentând starea sa de moment, nu departe de monitoarele folosite de controlorii de trafic pentru urmărirea avioanelor. De aici, datele senzoriale se îndreaptă către o regiune misterioasă numită talamus, care ne aduce la următorul nostru subiect.

Talamusul

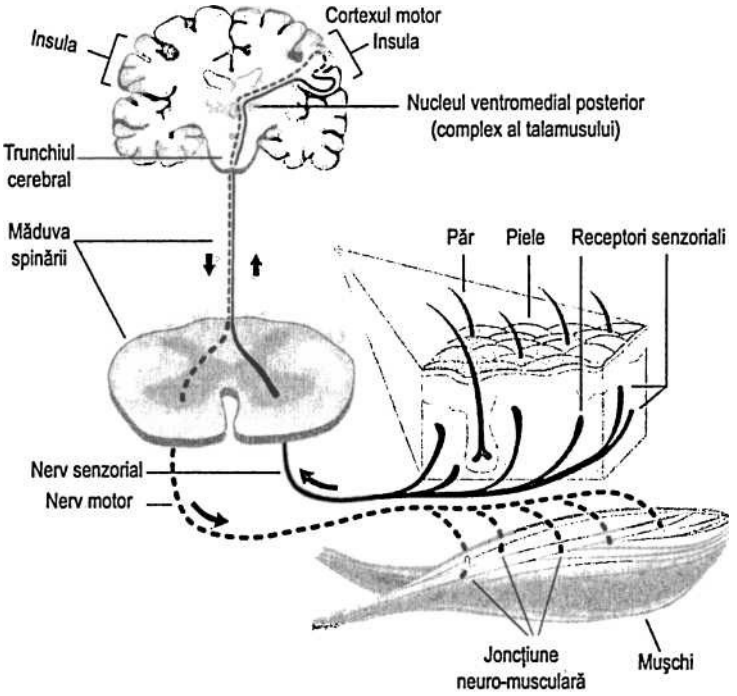
Toată lumea ştie ce înseamnă atenţia. Înseamnă preluarea de către minte, în formă limpede şi vie, a unuia din mai multe subiecte simultan posibile sau şiruri de gânduri. Focalizarea, concentrarea, conştientizarea (în de esenţa ei. Presupune retragerea de la unele obiecte pentru aprecierea eficientă a altora.

— William James

De la creierul mijlociu, informaţia senzorială este transmisă printr-o regiune de mărimea unei nuci, numită nucleu ventromedial posterior (VMpo) al talamusului, care computează reacţiile complexe la stări ale corpului, cum ar fi „E un gust îngrozitor”, „Ce duhoare” sau „Pata aceea de lumină e stimulatoare”. Informaţia procesată din ce în ce mai mult sfârşeşte în două regiuni ale neocortexului, numite insula. Aceste structuri, de dimensiunile unor degete mici, sunt localizate în partea stângă şi în cea dreaptă a neocortexului. Doctorul Arthur Craig de la Barrow Neurological Institute din Phoenix numeşte VMpo şi cele două regiuni ale insulei „sistem care reprezintă eul material”.4

Printre celelalte funcţii ale sale, talamusul este considerat poarta prin care informaţiile preprocesate ajung în neocortex. În plus faţă de informaţia tactilă care trece prin nucleul ventromedial posterior, informaţia procesată de la nervul optic (care, după cum am subliniat anterior, a fost deja substanţial transformată) este transmisă într-o regiune a talamusului numită nucleu geniculat lateral, care o trimite apoi la regiunea V1 a neocortexului. Informaţia de la receptorul auditiv trece

prin nucleul geniculat medial al talamusului, îndreptându-se către regiunile auditive prime ale neocortexului. Toate datele



Regiunea căilor senzorial-tactile la nivelul creierului

noastre senzoriale (cu excepţia, s-ar părea, a sistemului olfactiv, care foloseşte bulbii olfactivi) trece prin regiuni specifice ale talamusului.

Cel mai important rol al talamusului, însă, este comunicarea continuă cu neocortexul. Recognizatorii de modele de la nivelul neocortexului trimit rezultate provizorii la talamus şi primesc răspunsuri în principal folosind semnale reciproce atât excitatoare, cât şi inhibitoare, de la stratul VI al fiecărui recognizator. Reţine, cititorule, că nu este vorba de mesaje fără fir, aşadar trebuie să existe un volum extraordinar de legături materiale (sub formă de axoni) între toate regiunile neocortexului şi talamus. Gândeşte-te la volumul de spaţiu (în termeni de masă fizică de conexiuni necesare) pentru sutele de milioane de recognizatori de modele de la nivelul

neocortexului, care trebuie să se consulte permanent cu talamusul.5

Aşadar, despre ce vorbesc sutele de milioane de recognizatori de modele ai neocortexului cu talamusul? Aparent, este o conversaţie importantă, deoarece leziunile profunde bilaterale ale principalelor regiuni ale talamusului pot să conducă la o stare prelungită de pierdere a cunoştinţei. O persoană care prezintă leziuni ale talamusului mai poate să aibă activitate la nivelul neocortexului, în sensul că gândirea automată prin asociere mai poate încă funcţiona. Dar gândirea direcţionată – genul care ne face să ne ducem la culcare, să urcăm în maşină şi să ne aşezăm la birou la slujbă – nu funcţionează în absenţa talamusului. Într-un caz celebru, Karen Ann Quinlan, în vârstă de douăzeci şi unu de ani, a suferit un infarct miocardic şi manifesta insuficienţă respiratorie, aşa că a rămas într-o stare apatică, aparent vegetativă, timp de zece ani. După moarte, autopsia a dezvăluit faptul că neocortexul ei era normal, dar talamusul fusese distrus.

Pentru a-şi îndeplini rolul în privinţa abilităţii noastre de a ne direcţiona atenţia, talamusul se bazează pe cunoaşterea structurată conţinută de neocortex. El poate parcurge o listă (stocată în neocortex), permiţându-ne să urmărim un şir de gânduri sau să respectăm un plan de acţiune. Aparent, suntem capabili să avem, în acelaşi timp, în jur de patru subiecte active în memoria de lucru, câte două în fiecare emisferă, potrivit recentelor cercetări ale specialiştilor în neuroştiinţă de la MIT Picower Institute for Learning and Memory.6 Problema dacă talamusul este responsabil de neocortex sau invers e departe de a fi clarificată, însă nu suntem în măsură să funcţionăm în lipsa ambelor.

Hipocampul

Fiecare emisferă cerebrală are un hipocamp, o mică regiune care seamănă cu un căluţ de mare, ghemuit medial în lobul temporal. Principala sa funcţie este să-şi

amintească evenimentele noi. De vreme ce informaţia senzorială trece prin neocortex, de el depinde să stabilească dacă o experienţă este nouă pentru a o prezenta hipocampului. Şi face acest lucru fie nerecunoscând un anumit set de caracteristici (de exemplu, un chip nou), fie realizând faptul că o situaţie altminteri familiară are într-un moment actual atribute unice (cum ar fi acela că soţia ta are o mustaţă falsă).

Hipocampul este capabil să reţină asemenea situaţii, cu toate că s-ar părea că face acest lucru în primul rând prin anumite indicatoare din neocortex. Prin urmare, amintirile din hipocamp sunt stocate şi ca modele de nivel inferior, care au fost anterior recunoscute şi depozitate în neocortex.

Pentru ca animalele care nu au neocortex să poată modula experienţele senzoriale, hipocampul pur şi simplu memorează informaţia de la simţuri, deşi aceasta e trecută deja prin preprocesarea senzorială (de exemplu, transformarea efectuată de nervul optic). Deşi hipocampul se foloseşte de neocortex (dacă creierul respectiv beneficiază de aşa ceva) ca de un carneţel de notiţe, memoria sa (de indicatoare din neocortex) nu este în mod inerent ierarhică. Animalele lipsite de neocortex pot, prin urmare, să memoreze diverse lucruri folosindu-şi hipocampul, dar amintirile lor nu sunt ierarhice.

Capacitatea hipocampului este limitată, aşadar memorarea efectuată de el este de scurtă durată. Hipocampul transferă o anumită succesiune de modele din memoria sa de scurtă durată în memoria ierarhică de lungă durată a neocortexului, repetându-i acestuia, iar şi iar, respectiva succesiune de memorie. Prin urmare, avem nevoie de hipocamp pentru a memora lucruri şi abilităţi noi (deşi abilităţile strict motorii par să folosească un mecanism diferit). O persoană care a suferit tulburări la ambele copii ale hipocampului îşi păstrează amintirile existente, dar nu-şi mai poate crea altele noi.

Theodore Berger, specialist în neuroştiinţă la

Universitatea South California, împreună cu colaboratorii săi, a realizat un model al hipocampului de şobolan şi a experimentat cu succes implantarea unuia nou. În cadrul unui studiu ale cărui rezultate au fost comunicate în anul 2011, oamenii de ştiinţă de la Universitatea South California au blocat la şobolani, cu ajutorul medicamentelor, anumite comportamente învăţate. Folosindu-se de un hipocamp artificial, şobolanii au fost repede capabili să reînveţe comportamentul, „închizi comutatorul şi şobolanul îşi aminteşte. îl deschizi şi şobolanul uită", scria Berger referindu-se la posibilitatea lui de a controla de la distanţă implantul neuronal. În cadrul unui alt experiment, oamenii de ştiinţă au permis hipocampului lor artificial să funcţioneze în paralel cu cel natural al şobolanului. Rezultatul a fost acela că s-a consolidat abilitatea şobolanilor de a învăţa noi abilităţi. „Aceste studii integrate de modelare experimentală arată, pentru prima oară", explica Berger, „că... o proteză neuronală capabilă de identificarea şi manipularea în timp real a procesului de codificare poate să refacă şi chiar să stimuleze procesele cognitive mnemonice".7 Hipocampul este una dintre primele regiuni afectate de boala Alzheimer, deci unul dintre scopurile acestor cercetări îl reprezintă punerea la punct a unui implant neuronal pentru oameni, implant care va rezolva această primă fază a leziunilor provocate de boală.

Cerebelul

Există două modalităţi la care poţi apela pentru a încerca să prinzi o minge în zbor. Poţi să rezolvi ecuaţiile diferenţiale simultane complexe care controlează mişcarea mingii, dar şi alte ecuaţii care-ţi guvernează unghiul particular sub care vezi mingea, iar apoi să computezi şi mai multe ecuaţii pentru mişcarea corpului, braţului şi mâinii pentru a te afla în locul potrivit la momentul potrivit.

Însă nu aceasta este abordarea pe care o adoptă creierul tău. În esenţă, el simplifică problema, reducând o

mulţime de ecuaţii la un simplu model al orientării, apreciind locul în care pare să se afle mingea în câmpul vizual şi rapiditatea cu care se deplasează aceasta. Creierul face acelaşi lucru şi cu mâna ta, realizând predicţii liniare ale poziţiei aparente a mingii în câmpul vizual şi ale mişcărilor mâinii. Scopul, desigur, este acela ca el să se asigure că cele două elemente se întâlnesc în acelaşi punct din spaţiu şi din timp. Dacă mingea pare să cadă prea repede şi mâna pare să se mişte prea încet, creierul îţi face mâna să se mişte mai repede, astfel încât traiectoriile să coincidă. Această soluţie gen „nodul gordian" pentru ceea ce altminteri ar reprezenta o problemă insolvabilă matematic se numeşte funcţie de bază şi e îndeplinită de cerebel, o regiune cu forma unui bob de fasole şi de mărimea unei mingi de baseball, situată deasupra trunchiului cerebral.6

Cerebelul este o regiune a creierului vechi care a controlat cândva practic toate mişcările hominizilor. El încă mai conţine jumătate dintre neuronii creierului, deşi cei mai mulţi sunt relativ mici, astfel încât regiunea reprezintă numai în jur de 10 la sută din greutatea creierului. Pe de altă parte, cerebelul este o dovadă de masivă repetiţie în construcţia creierului. Există puţine informaţii în genom cu privire la construcţia lui, structura sa fiind un model de mai mulţi neuroni care se repetă de miliarde de ori. La fel ca în cazul neocortexului, el prezintă o uniformitate a întregii sale structuri.9

Cea mai mare parte a funcţiilor de control asupra activităţii musculare a fost preluată de neocortex, folosindu-se aceiaşi algoritmi de recunoaştere a modelelor care sunt utilizaţi şi pentru percepţie şi cogniţie. În cazul mişcării, la funcţia neocortexului ne putem referi mai adecvat prin sintagma „implementarea modelului". Într-adevăr, neocortexul se foloseşte de memoria din cerebel pentru înregistrarea scenariilor delicate ale mişcărilor de exemplu, semnătura şi anumite înfloriturі de expresie artistică, bunăoară în muzică şi dans. Studiile asupra

rolului cerebelului în timpul învăţării scrierii de mână de către copii dezvăluie faptul că celulele Purkinje de la nivelul cerebelului iau de fapt mostre de mişcări, fiecare celulă fiind sensibilă la o anumită mostră.10 Deoarece mişcările noastre, în cea mai mare parte a lor, sunt controlate de neocortex, mulţi oameni se pot descurca bine în condiţiile unor dizabilităţi evidente relativ modeste, ba chiar şi în cazul lezării cerebelului, doar că mişcările pot deveni mai puţin graţioase.

Neocortexul mai poate recurge la cerebel şi pentru a-i folosi abilitatea de computare a funcţiilor de bază în timp real, în vederea anticipării rezultatelor acţiunilor pe care le avem în vedere, dar pe care nu le-am realizat încă (şi e posibil să nici nu le efectuăm vreodată), precum şi a acţiunilor sau a posibilelor acţiuni ale celorlalţi. Este încă o dovadă a existenţei predictorilor liniari înnăscuţi incluşi în creier.

S-au înregistrat progrese substanţiale în simularea cerebelului sub aspectul capacităţii sale de a răspunde dinamic la indicii senzoriale, folosind funcţiile de bază pe care le-am descris anterior atât în simulările inversate (bazate pe modele biochimice), cât şi. În simulările directe (bazate pe modele matematice ale operării la nivelul cerebelului a fiecărei unităţi repetate).11

Plăcere şi frică

Frica este principala sursă de superstiţii şi una dintre principalele surse de cruzime. Înfrângerea fricii este începutul înţelepciunii.

— Bertrand. Russell

Simte frica şi alung-o.

— Susan Jeffers

Dacă neocortexul este bun la rezolvarea problemelor, atunci care-i principala problemă pe care vrem să o rezolvăm? Problema pe care evoluţia a încercat întotdeauna s-o soluţioneze o reprezintă supravieţuirea speciilor. Aceasta se traduce prin supravieţuirea individului şi fiecare dintre noi îşi foloseşte propriul

neocortex ca să interpreteze acest lucru într-o multitudine de moduri. Pentru a supravieţui, animalele sunt nevoite să-şi procure următoarea masă şi, concomitent, să evite să devină masa altui animal. De asemenea, trebuie să se reproducă. Creierul timpuriu şi-a dezvoltat sisteme de frică şi plăcere care să recompenseze îndeplinirea acestor nevoi fundamentale, laolaltă cu comportamente care le facilitau. Pe măsură ce mediul şi speciile concurente s-au modificat, evoluţia biologică a făcut schimbările corespunzătoare. Odată cu apariţia gândirii ierarhizate, satisfacţia pornirilor fundamentale a devenit mai complexă, fiind de-acum supusă vastului complex de idei din idei. Însă, în ciuda considerabilei modulări de către neocortex, creierul vechi este în continuare viu şi încă ne mai motivează cu plăcerea şi cu frică.

O regiune asociată cu plăcerea este nucleus accumbens. În celebrele experimente realizate în anii 1950, şobolanii s-au arătat capabili să îşi stimuleze direct această mică regiune (prin apăsarea unei pedale care activa electrozii implantaţi), preferând această activitate oricărei alteia, inclusiv raportului sexual sau hrănirii, iar în cele din urmă epuizându-se şi înfometându-se până la moarte.12 La om, mai există şi alte regiuni implicate în plăcere, cum ar fi nucleus pallidum ventral şi, desigur, neocortexul însuşi.

Plăcerea este provocată şi de compuşi chimici cum ar fi dopamina şi serotonina. Nu intră în scopurile acestei cărţi discutarea detaliată a acestor sisteme, dar e important să recunoaştem că am moştenit mecanismele cu pricina de la verişoarele noastre premamifere. Menirea neocortexului este aceea de a ne permite să fim stăpânii plăcerii şi fricii, nu sclavii lor. Cum însă suntem adeseori supuşi unor comportamente de dependenţă, neocortexul nu are întotdeauna succes în acest sens. Dopamina, mai ales, este un neurotransmiţător implicat în trăirea plăcerii. Când ni se întâmplă ceva plăcut – câştigăm la loterie, obţinem recunoaşterea colegilor, primim o îmbrăţişare de

la o persoană iubită sau chiar înregistrăm realizări subtile, cum ar fi să-l facem pe un prieten să râdă la o glumă –, trăim experienţa unei descărcări de dopamină. Uneori şi noi, asemenea şobolanilor care au murit stimulându-şi excesiv nucleus accumbens, folosim o scurtătură pentru obţinerea acestor accese de plăcere, ceea ce nu e întotdeauna o idee bună.

Jocurile de noroc, de exemplu, pot descărca dopamină cel puţin atunci când câştigi, dar asta depinde de inerenta lipsă de predictibilitate a situaţiei. Jocurile de noroc pot sluji scopul descărcării dopaminei pentru un timp, dar, dat fiind faptul că şansele sunt intenţionat create pentru a lucra împotriva ta (altminteri nu ar funcţiona modelul de afacere al unui cazinou), pot deveni ruinătoare ca strategie regulată. Pericole similare sunt asociate cu oricare alt comportament de dependenţă. O anumită mutaţie a genei receptorului de dopamină D_2 provoacă sentimente de plăcere deosebit de puternice la primele experienţe cu substanţe adictive şi cauzează comportamente ce dau dependenţă, dar, după cum este bine ştiut (însă nu întotdeauna ţinut minte), capacitatea acestor substanţe de a produce plăcere la utilizările ulterioare de obicei se reduce. O altă mutaţie genetică rezidă în faptul că unele persoane nu beneficiază de descărcarea unor cantităţi normale de dopamină în urma realizărilor cotidiene, ceea ce poate de asemenea să conducă la căutarea unor experienţe stimulatoare prin activităţi care dau dependenţă. Minoritatea celor care manifestă aceste tendinţe genetice de dependenţă creează o enormă problemă socială şi medicală. Chiar şi persoanele care reuşesc să evite comportamentele sever dependente se luptă cu echilibrarea recompenselor la descărcările de dopamină.

Serotonina este un neurotransmiţător care joacă un rol major în reglarea stărilor de spirit. La niveluri mai înalte, este asociată cu sentimente de bine şi mulţumire. Serotonina are şi alte funcţii, inclusiv modularea tăriei

sinaptice, a apetitului, somnului, dorinţei sexuale şi digestiei. Medicamentele antidepresive, cum ar fi inhibitorii selectivi de refacere a serotoninei (care tind să sporească nivelul de serotonină disponibilă pentru receptori), tind să aibă efecte cu bătaie lungă – nu toate dezirabile (cum ar fi suprimarea libidoului). Spre deosebire de acţiunile de la nivelul neocortexului, unde recunoaşterea modelelor şi activarea axonilor afectează concomitent doar un număr redus de circuite neocorticale, antidepresivele afectează regiuni întinse ale creierului sau chiar întregul sistem nervos.

Fiecare emisferă a creierului uman are o amigdală, care este o regiune în formă de migdală, conţinând mai mulţi lobi mici. Amígdala e şi ea parte a creierului vechi şi e implicată în procesarea mai multor tipuri de răspunsuri emoţionale, cel mai notabil dintre ele fiind frica. La animalele premamifere, anumiţi stimuli preprogramaţi care semnalează pericolul se transmit direct amigdalei, care, la rândul ei, declanşează mecanismul „Luptă sau fugi”. La om, actualmente, amígdala este importantă pentru transmiterea percepţiei de pericol la neocortex. Un comentariu negativ al şefului, de exemplu, poate declanşa un asemenea răspuns prin generarea fricii de pierdere a slujbei (sau poate nu, dacă ai încredere într-un plan B). Din momentul în care amígdala decide că te pândeşte un pericol, survine o succesiune ancestrală de evenimente. Amígdala îi transmite glandei pituitare să descarce un hormon denumit ACTH (adrenocorticotropină). Acesta, la rândul său, stimulează descărcarea în sânge a hormonului de stres cortizol, localizat în glandele suprarenale, iar prezenţa acestuia se concretizează în asigurarea unei energii mai mari la nivelul sistemului muscular şi nervos. Glandele suprarenale secretă, de asemenea, adrenalină şi noradrenalină, care inhibă sistemele digestiv, imunitar şi reproducător (imaginându-şi pă nu sunt procese de mare prioritate într-o situaţie de urgenţă). Cresc în schimb nivelurile tensiunii arteriale, glicemiei, colesterolului şi

fibrinogenului (care accelerează coagularea sângelui). Bătăile inimii şi respiraţia se înteţesc. Chiar şi pupilele se dilată, astfel încât să existe o mai bună acuitate vizuală pentru observarea inamicului sau pentru identificarea căii de scăpare. Toate acestea sunt foarte utile dacă întâlneşti pe neaşteptate un pericol, cum ar fi un prădător. Este foarte bine cunoscut faptul că, în lumea zilelor noastre, activarea cronică a mecanismului „Luptă sau fugi" poate conduce la afecţiuni permanente sub forma hipertensiunii, hipercolesterolemiei ş.a.

Sistemul global al nivelurilor neurotransmiţătorilor cum ar fi serotonina şi al hormonilor cum ar fi dopamina este foarte complicat şi am putea să ocupăm tot restul cărţii de faţă cu acest subiect (cum s-a şi întâmplat cu foarte multe cărţi), dar merită să subliniez că lăţimea de bandă a informaţiei (rata de procesare a informaţiei) din acest sistem e foarte redusă comparativ cu lăţimea de bandă a neocortexului. Există un număr limitat de substanţe implicate, iar nivelurile acestor compuşi chimici tind să se modifice încet şi sunt relativ universale în tot creierul, spre deosebire de neocortex, care e format din sute de miliarde de celule ce se pot schimba rapid.

Este corect să spun că experienţele noastre emoţionale se produc atât la nivelul creierului vechi, cât şi la cel al creierului nou. Gândirea are loc în creierul nou (neocortexul), dar sentimentele apar în ambele. Orice imitare a comportamentului uman trebuie, prin urmare, să conţină modele ale ambelor. Pe de altă parte, dacă urmărim numai inteligenţa cognitivă a omului, neocortexul este suficient. Putem înlocui creierul vechi cu o motivaţie mai directă a unui neocortex nonbiologic pentru obţinerea scopurilor pe care i le atribuim. De exemplu, în cazul lui Watson, scopul a fost formulat simplu: Dă răspunsurile corecte la întrebările de la *Jeopardy*! (cu toate că acestea erau modulate şi mai mult de un program care înţelegea cum se joacă *Jeopardy*!). În cazul noului sistem Watson, care este pus la punct, în colaborare, de Nuance şi IBM

pentru achiziţia de cunoştinţe medicale, scopul este ca acesta să ajute la tratarea afecţiunilor omului. Sistemele viitoare vor putea să aibă scopuri cum ar fi vindecarea propriu-zisă a bolilor şi eliminarea sărăciei. În mare măsură, lupta plăcere-frică este deja străveche pentru om, creierul vechi evoluând cu mult înainte chiar şi de apariţia societăţii umane primitive; într-adevăr, în marea sa parte, acest creier este reptilian.

În creierul omului se dă o luptă continuă: aparţine controlul creierului vechi sau creierului nou? Creierul vechi încearcă să facă agenda cu controlul experienţelor de plăcere şi frică, în vreme ce creierul nou se străduieşte din răsputeri să înţeleagă algoritmii relativ primitivi ai creierului vechi şi să îi manipuleze după propriile interese. Reţine că amigdala este incapabilă să evalueze singură pericolul – creierul omului se bazează pe neocortex pentru a judeca un asemenea lucru. Persoana aceea e prieten sau duşman, o persoană iubită sau o ameninţare? Numai neocortexul poate să decidă.

În măsura în care nu suntem implicaţi direct într-o luptă pe viaţă şi pe moarte şi nu vânăm ca să ne hrănim, am reuşit să ne sublimăm cel puţin parţial pornirile ancestrale în sensul unor acţiuni mai temerare. Pe această notă, vom discuta creativitatea şi iubirea în capitolul care urmează.

Capitolul 6
ABILITĂŢILE TRANSCENDENTE

Aceasta este religia mea simplă. Nu e nevoie de temple; nu e nevoie de o filosofie complicată. Propriul creier şi propria inimă ne sunt templul; filosofia este generozitatea.

— Dalai Lama

Mâna mi se mişcă pentru că anumite forje – electrice, magnetice sau orice altă „forţă nervoasă” – îmi sunt imprimate în creier. Forţei nervoase depozitate în creier probabil că i s-ar putea da de urmă, dacă ştiinţa ar fi

completă, în forţele chimice puse la dispoziţia creierului de către sânge şi e obţinută, în cele din urmă, din alimentele pe care le mănânc şi din aerul pe care îl respir.

— Lewis Carroll

Tot în neocortex îşi au sediul şi gândurile noastre emoţionale, dar sunt influenţate de părţi ale creierului care merg de la regiuni ale creierului vechi, precum amigdala, până la structuri ale creierului recente din punct de vedere evolutiv, cum ar fi neuronii fusiformi, care par să joace un rol-cheie în emoţiile de ordin superior. Spre deosebire de structurile recursive regulate şi logice care se găsesc în cortexul cerebral, neuronii fusiformi prezintă contururi şi conexiuni foarte neregulate. Sunt cei mai mari neuroni din creierul uman, acoperind toată întinderea lui, şi sunt profund interconectaţi, cu sute de mii de legături care unesc unele de altele diferite porţiuni ale neocortexului.

După cum am menţionat anterior, insula ajută la procesarea semnalelor senzoriale, dar joacă un rol-cheie şi în emoţiile de ordin superior. Aceasta este regiunea în care îşi au originea neuronii fusiformi. Scanările pe bază de rezonanţă magnetică în condiţii funcţionale (RMN) au dezvăluit faptul că aceste celule sunt deosebit de active atunci când persoana se confruntă cu sentimente cum ar fi iubirea, furia, tristeţea şi cu dorinţa sexuală. Situaţiile care le activează intens le includ şi pe cele în care o femeie îşi priveşte partenerul sau îşi aude copilul plângând.

Neuronii fusiformi au filamente neuronale lungi, denumite dendrite apicale, care sunt capabile să se conecteze cu regiuni îndepărtate ale neo. cortexului. O asemenea interconectare „profundă”, prin care anumiţi neuroni asigură conectarea la distanţe de mai multe regiuni, reprezintă o caracteristică tot mai evidentă pe măsură ce urcăm pe scara evoluţiei. Nu este surprinzător faptul că neuronii fusiformi, atât de implicaţi în emoţii şi judecăţi morale, au acest nivel de conectare, dată fiind capacitatea reacţiilor emoţionale de nivel superior de a

atinge diverse aspecte şi gânduri.

Datorită legăturilor lor cu numeroase regiuni ale creierului, emoţiile de nivel superior pe care le procesează neuronii fusiformi sunt afectate de toate regiunile noastre legate de percepţie şi cogniţie. Este important să subliniez faptul că aceste celule nu realizează o rezolvare raţională a problemelor, motiv pentru care nici nu avem un control raţional asupra răspunsurilor noastre la muzică sau la momentele când ne îndrăgostim. Pe de altă parte, restul creierului e puternic angajat în încercarea de a desluşi misterioasele noastre emoţii de ordin superior.

Există relativ puţini neuroni fusiformi: doar în jur de 80.000, din care aproximativ 45.000 în emisfera dreaptă şi 35.000 în emisfera stângă. Această disparitate explică parţial percepţia potrivit căreia inteligenţa emoţională este domeniul emisferei drepte a creierului, deşi disproporţia e modestă. Gorilele au în jur de 16.000 de asemenea celule, cimpanzeul pitic are cam 2.100, iar cimpanzeii – în jur de 1.800. De la alte mamifere, ele lipsesc cu desăvârşire.

Antropologii cred că neuronii fusiformi şi-au făcut apariţia în urmă cu 10 până la 15 milioane de ani, la deocamdată nedescoperitul strămoş al maimuţelor şi al omului şi că au crescut rapid numeric acum aproximativ 100.000 de ani. Fapt interesant, neuronii fusiformi nu există la nou-născutul uman şi încep să apară abia în jurul vârstei de patru luni, iar numărul lor creşte semnificativ între vârstele de unu şi trei ani. Capacitatea copiilor de a face faţă temelor morale şi de a percepe emoţii de ordin superior cum ar fi iubirea se dezvoltă exact în aceeaşi perioadă.

Aptitudinile

Wolfgang Amadeus Mozart (1756 – 1791) a scris un menuet pe când avea cinci ani. La vârsta de şase ani, a cântat în faţa împărătesei Maria

Tereza, la curtea imperială de la Viena. A compus şase sute de piese, inclusiv patruzeci şi una de simfoniunaintea morţii sale survenite la vârsta de treizeci

şi cinci de ani şi este considerat îndeobşte cel mai mare compozitor european de muzică clasică. S-ar putea spune că a avut aptitudini pentru muzică.

Aşadar, ce înseamnă asta în contextul Teoriei minţii bazate pe recunoaşterea modelelor? în mod evident, ceea ce considerăm că reprezintă o aptitudine este produsul cultivării – cu alte cuvinte, al influenţelor mediului şi al altor oameni. Mozart s-a născut într-o familie muzicală. Tatăl lui, Leopold, era compozitor şi *kapellmeister* (conducător muzical, literal) al orchestrei curţii arhiepiscopului de Salzburg. Tânărul Mozart a fost scufundat în muzică şi tatăl lui a început să-l înveţe vioara şi clavirul (un instrument cu claviaturi) de la vârsta de trei ani.

Pe de altă parte însă, numai influenţele mediului nu pot să explice complet geniul lui Mozart. Există aici, în mod evident, şi o componentă naturală. Ce formă îmbracă aceasta? După cum am scris în Capitolul 4, diferite regiuni ale neocortexului au ajuns să fie optimizate (prin evoluţia biologică) pentru anumite tipuri de modele. Cu toate că algoritmul de bază al recunoaşterii modelelor este uniform în tot neocortexul, dat fiind faptul că anumite tipuri de modele tind să acopere anumite regiuni (chipurile la nivelul girusului fusiform, de exemplu), aceste regiuni ajung să fie mai performante la procesarea modelelor asociate.

În orice caz, există numeroşi parametri care guvernează modul cum este aplicat propriu-zis algoritmul la nivelul fiecărui modul. De exemplu, cât de mare trebuie să fie asemănarea pentru ca modelul să fie recunoscut? Cum se modifică acest prag dacă un modul de nivel superior trimite un semnal potrivit căruia este „aşteptat” modelul său? Cum sunt apreciaţi parametrii de mărime? Aceşti factori şi alţii au fost stabiliţi diferit în regiuni diferite, pentru a reprezenta un avantaj pentru anumite tipuri de modele. În cercetările noastre pe metode similare în domeniul inteligenţei artificiale, am observat acelaşi

fenomen şi am folosit simulări ale evoluţiei pentru optimizarea acestor parametri.

Dacă anumite regiuni pot fi optimizate pentru diferite tipuri de modele, rezultă că creierele individuale variază şi ele, prin capacitatea de învăţare, de recunoaştere şi de creare a anumitor tipuri de modele. De exemplu, un creier poate avea aptitudini înnăscute pentru muzică, fiind capabil într-o măsură mai mare să recunoască modelele ritmice, ori poate fi mai înclinat să înţeleagă aranjamentele geometrice ale armoniilor. Fenomenul înălţimii perfecte (capacitatea de a recunoaşte şi a reproduce o înălţime fără un reper exterior), corelat cu talentul muzical, pare să aibă o bază genetică, deşi capacitatea aceasta trebuie cultivată, ceea ce înseamnă că este foarte posibil să fie o combinaţie de natură şi cultură. Baza genetică a înălţimii perfecte e posibil să rezide în afara neocortexului, în preprocesarea informaţiei auditive, pe când aspectul învăţat rezidă în neocortex.

Există şi alte abilităţi care contribuie la gradul de competenţă, indiferent dacă este vorba de genul rutinier sau de geniul legendar. Abilităţile neocorticale – de exemplu, capacitatea neocortexului de a controla semnalele de frică pe care le generează amigdala (când se confruntă cu dezaprobarea) – joacă un rol semnificativ, la fel ca alte atribute, cum ar fi încrederea, abilităţile organizaţionale şi capacitatea de a-i influenţa pe alţii. O abilitate foarte importantă pe care am scos-o în evidenţă mai devreme este curajul de a urmări idei care merg împotriva preceptelor ortodoxe. Invariabil, oamenii pe care îi considerăm genii şi-au luat în considerare propriile experimente mentale în moduri care iniţial nu au fost înţelese sau apreciate de către confraţii lor. Deşi Mozart şi-a câştigat recunoaşterea în cursul vieţii, cea mai mare parte a adulaţiei a venit abia mai târziu. A murit sărac, a fost înmormântat într-o groapă comună, iar la funeraliile lui au participat doar alţi doi muzicieni.

Creativitatea

Creativitatea este un drog fără care nu pot să trăiesc.

— Cecil B. DeMille

Problema nu este. niciodată cum să-ţi bagi în minte idei noi, inovatoare, ci cum să le scoţi pe cele vechi. Fiecare minte este o clădire plină de mobilă veche. Fă-ţi curăţenie într-un colţ al minţii şi instantaneu îl va umple creativitatea.

— Dee Hock

Umanitatea poate fi cât se poate de rece faţă de cei ai căror ochi văd lumea altfel.

— Eric A. Burns

Creativitatea poate rezolva aproape orice problemă. Actul creator, înfrângerea obişnuinţei de către originalitate, biruieşte orice.

— George Lois

Un aspect-cheie al creativităţii este procesul de găsire a marilor metafore – simboluri care reprezintă altceva. Neocortexul e o mare maşină de produs metafore, fapt ce explică de ce suntem o specie extraordinar de creativă. Oricare dintre cele aproximativ 300 de milioane de recognizatori de modele din neocortexul nostru recunoaşte, defineşte un model şi îi dă un nume. Asta, în cazul modulelor neocorticale de recunoaştere a modelelor, înseamnă pur şi simplu că un axon iese din recognizatorul de modele şi emite un impuls atunci când găseşte modelul respectiv. La rândul lui, acest simbol devine parte a altui model. Fiecare dintre aceste modele este, în esenţă, o metaforă. Recognizatorii pot să emită impulsuri şi de 100 de ori pe secundă, aşadar avem potenţialul de recunoaştere a unui număr de 30 de miliarde de metafore pe secundă. Desigur, nu toate modulele emit semnale la fiecare ciclu – însă este corect să se spună că, într-adevăr, recunoaştem milioane de metafore pe secundă.

Desigur, unele metafore sunt mai semnificative decât altele. Darwin şi-a dat seama că intuiţia lui Charles Lyell, potrivit căreia modificările extrem de graduale provocate de o picătură de apă pot săpa canioane imense, constituie

o metaforă sugestivă a modului în care o picătură de mici modificări evolutive, pe parcursul miilor de generaţii, poate produce modificările imense în diferenţierea speciilor. Experimentele mintale, cum ar fi cel pe care l-a folosit Einstein pentru a pune în lumină adevărata semnificaţie a experimentului Michelson-Morley, sunt toată metafore, în sensul că reprezintă un „lucru considerat reprezentativ sau simbolic pentru altceva”, ca să citez o definiţie de dicţionar.

Vezi vreo metaforă în *Sonetul 73* de Shakespeare?

Acea vreme-a anului ce-n mine ai vedea-o
Când frunze galbene, ori niciuna, ori câteva, mai stau
Pe cele crengi ce tremură de frig
Strane pustii năruite, în care odinioară păsări cântau.
În mine vezi amurgul unei asemenea zile
Cum dup-apus se pierde-n miazănoapte.
Pe care-ncet-încet noaptea neagră trebuie să-l răpească.
A doua faţ-a morţii ce pecetluieşte totul în odihnă, în mine vezi sclipirile unui asemenea foc Ce pe cenuşa tinereţii trebuie să ardă la *pe-un pat al morţii pe care trebuie să se sfârşească înghiţit de ceea ce a fost hrănit.*
Asta vezi şi iubirea ţi-o întăreşte.
Mult să iubeşti ce-aici trebuie să laşi.

În acest sonet, poetul foloseşte metafore extinse pentru a-şi descrie înaintarea în vârstă. Vârsta lui e precum toamna târzie, „când frunze galbene, ori niciuna, ori câteva, mai stau”. Vremea e rece şi păsările nu mai pot rămâne pe ramuri, pe care poetul le numeşte „strane pustii năruite”. Vârsta lui este asemenea apusului: „cum dup-apus se pierde-n miazănoapte/ pe care-ncet-încet noaptea neagră trebuie să-l răpească”. El este restul unui foc „ce pe cenuşa tinereţii trebuie să ardă”. Pe bună dreptate, întregul limbaj este în ultimă instanţă o metaforă, cu expresii unele mai memorabile decât altele.

Găsirea unei metafore reprezintă procesul de recunoaştere a unui model în ciuda diferenţelor de detaliu

şi context – o activitate pe care o efectuăm în mod curent, în fiecare moment al vieţii noastre. Salturile metaforice pe care le considerăm semnificative tind să aibă loc în spaţiile dintre diferite discipline. Pe de altă parte, acţionând împotriva forţei esenţiale a creativităţii, manifestăm tendinţa atotprezentă spre o tot mai profundă specializare în ştiinţe (şi în aproape toate celelalte domenii, de altfel). După cum a scris matematicianul american Norbert Wiener (1894 – 1964) în lucrarea sa fundamentală *Cibernetica*, publicată în anul în care m-am născut eu (1948):

Există domenii de cercetare ştiinţifică, după cum vom vedea în cuprinsul acestei cărţi, care au fost explorate din perspectivele diferite ale matematicilor pure, statisticii, ingineriei electrice şi neurofiziologiei; domenii în care absolut toate noţiunile capătă un nume diferit de la fiecare grup şi în care cercetări foarte importante au fost triplate sau cvadruplate, în timp ce alte cercetări importante sunt întârziate de lipsa dintr-un domeniu a unor rezultate care e posibil să fi devenit deja clasice într-un alt domeniu.

Aceste domenii de frontieră oferă cele mai bogate oportunităţi cercetătorului pregătit. În acelaşi timp, sunt cele mai refractare la tehnicile acceptate de atac 7 n masă şi de diviziune a muncii.

O tehnică pe care am folosit-o în propriile mele cercetări pentru combaterea specializării tot mai accentuate a fost să adun experţi pentru un proiect (de exemplu, cercetările mele de recunoaştere a vorbirii au implicat oameni de ştiinţă, lingvişti, psihoacusticieni şi experţi în recunoaşterea modelelor, fără să-i mai menţionez pe cercetătorii din domeniul computerului) şi să-i încurajez pe toţi să predea grupului tehnicile şi terminologia care le erau particulare. Pe urmă, ne-am descotorosit de toate aceste terminologii şi am conceput-o pe a noastră. Invariabil, am găsit metafore dintr-un domeniu care rezolvă probleme în altul.

Un şoarece care găseşte o cale de scăpare atunci

când ajunge faţă în faţă cu pisica din gospodărie – şi care poate face acest lucru chiar şi dacă situaţia este pe undeva diferită de tot ceea ce a trăit anterior – dă dovadă de creativitate. Propria noastră creativitate este cu câteva ordine de mărime mai înaltă decât cea a şoarecelui – şi implică mult mai multe niveluri de abstracţie – deoarece noi avem un cortex mult mai mare, capabil de niveluri mult mai mari de ierarhizare. Prin urmare, un mod de asigurare a unei creativităţi mult mai mari constă în asamblarea eficientă a unui neocortex mai mare.

O abordare a expansiunii neocortexului disponibil se poate face prin colaborarea mai multor oameni. Operaţia se realizează în mod curent prin comunicarea dintre oamenii reuniţi într-o comunitate pentru rezolvarea unei probleme. Recent, s-au întreprins eforturi pentru folosirea instrumentelor de colaborare online în scopul exploatării forţei colaborării în timp real, iar strategia a avut succes în matematică şi în alte domenii.1

Desigur, următorul pas va consta în extinderea neocortexului însuşi, prin echivalentul lui nonbiologic. Acesta va fi actul nostru suprem de creativitate: să creăm capacitatea de a fi creativi. Un neocortex nonbiologic ar fi, în primul şi în primul rând, mai rapid şi ar putea să caute repede genurile de metafore care i-au inspirat pe Darwin şi Einstein. Ar putea să exploreze sistematic toate zonele de suprapunere a frontierelor dintre domeniile noastre de cunoaştere în continuă expansiune.

Unii îşi exprimă preocuparea pentru ce se va întâmpla cu aceia care vor opta să nu se supună unei asemenea expansiuni a minţii. Eu aş scoate în evidenţă faptul că această inteligenţă suplimentară va rezida în esenţă în *nor* (reţeaua în expansiune exponenţială de computere la care ne conectăm prin comunicarea online), unde este stocată acum cea mai mare parte a inteligenţei maşinilor noastre. Când foloseşti un motor de căutare, când recunoşti vorbirea cu telefonul, când consulţi un asistent virtual cum ar fi Siri sau îţi foloseşti telefonul

pentru a traduce un termen într-o altă limbă, inteligenţa nu se află în aparatul în sine, ci în nor. Neocortexul nostru extins va fi şi el găzduit tot acolo. Dacă accesăm o asemenea inteligenţă extinsă prin conexiuni neuronale directe, cum facem acum – interacţionând cu ea prin intermediul dispozitivelor noastre –, operăm o distincţie arbitrară. Din punctul meu de vedere, vom deveni mai creativi prin această amplificare atotprezentă, indiferent dacă alegem să ne conectăm sau nu la inteligenţa extinsă a omenirii. Deja am redirecţionat o mare parte din memoria noastră personală, socială, istorică şi culturală către nor şi în cele din urmă vom face acelaşi lucru şi cu gândirea noastră ierarhică.

Realizarea lui Einstein a rezultat nu numai din aplicarea metaforelor prin experimentele minţii, ci şi din curajul lui de a crede în forţa acestor metafore. A fost capabil să părăsească explicaţiile tradiţionale care nu reuşiseră să-i satisfacă experimentele şi s-a arătat dispus să înfrunte ridiculizarea confraţilor lui la aflarea bizarelor explicaţii pe care le implicau metaforele lui. Aceste calităţi – credinţa în metafore şi curajul convingerilor – sunt unele dintre cele pe care ar trebui să fim capabili să le programăm în neocortexul nostru nonbiologic.

Iubirea

Limpezimea minţii înseamnă şi limpezimea pasiunii; acesta este motivul pentru care o minte mare şi limpede iubeşte arzător şi vede distinct ceea ce iubeşte.

— Blaise Pascal întotdeauna, iubirea are o anumită nebunie. Dar există întotdeauna şi o anumită raţiune în nebunie.

— Friedrich Nietzsche

Când vezi atât de multe în viaţă câte am văzut eu, nu mai subestimezi forţa iubirii obsesive.

— Albuş Dumbledore, în J.K. Rowling, *Harry Potter şi prinţul vitreg* întotdeauna îmi place o soluţie matematică bună la orice problemă de iubire.

— Michael Patrick King, în episodul intitulat *Du-mă*

la meci din *Sex and the City*

Dacă nu ai trăit cu adevărat şi personal experienţa iubirii extatice, fără îndoială că ai auzit de ea. Este corect să spun că o parte substanţială, dacă nu chiar majoritatea operelor de artă ale lumii – poveşti, romane, muzică, dans, picturi, emisiuni de televiziune şi filme – sunt inspirate de poveşti de iubire în stadiile ei incipiente.

Recent, şi ştiinţa a intrat în acest domeniu, iar acum suntem capabili să identificăm schimbările biochimice care se produc atunci când cineva se îndrăgosteşte. Se descarcă dopamină, care produce sentimente de fericire şi desfătare. Nivelul sanguin al norepinefrinei creşte foarte mult, ceea ce duce la intensificarea bătăilor inimii şi la o stare generală de exaltare. Aceşti compuşi chimici, împreună cu feniletilamina, produc jubilare, niveluri ridicate de energie, concentrarea atenţiei, pierderea poftei de mâncare şi o sete arzătoare de a poseda obiectul dorinţelor. Fapt interesant, cercetări recente de la University College din Londra arată, de asemenea, că nivelul serotoninei în sânge se reduce, situaţie similară cu cea care are loc în cazul tulburărilor obsesiv-compulsive, fapt ce corespunde naturii obsesive a iubirii incipiente.2 Nivelurile ridicate ale dopaminei şi norepinefrinei explică creşterea atenţiei pe termen scurt, euforia şi dorinţa de a trăi iubirea incipientă.

Faptul că aceste fenomene biochimice par similare cu cele ale sindromului „Luptă sau fugi" este adevărat, doar că aici alergăm către ceva sau cineva; pe bună dreptate, cinicul ar putea să spună: *către* pericol, nu *departe* de el. De asemenea, schimbările corespund pe deplin cu cele ale fazelor incipiente ale comportamentului dependent. Piesa muzicală în stilul roxy music *Love Is the Drog* este cât se poate de elocventă în descrierea acestei stări (deşi subiectul piesei îi constituie căutarea următoarei satisfaceri cu iubire). Studiile pe marginea experienţelor religioase extatice indică, de asemenea, aceleaşi fenomene fizice; se poate spune că persoana care trăieşte o

experienţă mistică se îndrăgosteşte de Dumnezeu sau de orice fel de legătură spirituală asupra căreia se concentrează.

În cazul iubirii romantice incipiente, estrogenii şi ţestosteronul joacă în mod sigur un rol în stabilirea pornirii sexuale, dar, dacă reproducerea sexuală ar fi singurul obiectiv evolutiv al iubirii, atunci aspectul romantic al procesului nu ar mai fi necesar. Aşa cum scria psihologul John William Money (1921 – 2006), „dorinţa e desfrânată, iubirea e lirică”.

Faza extatică a iubirii conduce la faza de ataşament şi în cele din urmă la o legătură pe termen lung. Există compuşi chimici care încurajează şi acest proces, inclusiv ocitocina şi vasopresina. Să ne gândim la două specii înrudite de şoareci de câmp: şoarecele de câmp din prerie şi şoarecele de câmp de la munte. Sunt specii cât se poate de asemănătoare, cu excepţia faptului că şoarecele de prerie are receptori pentru ocitocină şi vasopresină, în timp ce şoarecele de munte nu îi are. Şoarecele de prerie este cunoscut pentru relaţiile sale monogame pe toată durata vieţii, în timp ce şoarecele de munte recurge aproape exclusiv la relaţii de o singură noapte. În cazul şoarecilor, ocitocina şi vasopresina au un rol determinant în privinţa naturii vieţii lor sexuale.

t

Deşi aceşti compuşi chimici exercită influenţe şi asupra omului, neocortexul nostru a preluat comanda, la fel ca în toate celelalte lucruri pe care le facem. Şoarecii de câmp au şi ei neocortex, dar cu dimensiuni de timbru şi plat. El este, totuşi, suficient de mare pentru ca şoarecii să îşi găsească o pereche pentru toată viaţa (sau, în cazul şoarecilor de munte, cel puţin de-o noapte) şi să-şi manifeste celelalte caracteristici elementare de şoareci. Noi, oamenii, avem suficient de mult neocortex adiţional ca să ne angajăm în expresiile expansive „lirice” la care se referă Money.

Dintr-o perspectivă evoluţionistă, chiar şi iubirea

există. pentru satisfacerea nevoilor neocortexului. Dacă nu am fi avut neocortex, atunci dorinţa carnală ar fi fost arhisuficientă pentru a ne garanta reproducerea. Provocarea extatică a iubirii conduce la ataşament şi dragoste matură şi se soldează cu o legătură durabilă. Aceasta, la rândul ei, este făcută să asigure cel puţin probabilitatea unui mediu stabil pentru copii în timpul în care neocortexul lor trece prin procesul critic de învăţare, necesar pentru a deveni adulţi responsabili şi capabili. Învăţarea într-un mediu bogat este o parte inerentă a metodei neocortexului. Aceleaşi mecanisme ale ocitocinei şi vasopresinei joacă un rol-cheie şi în stabilirea unei legături esenţiale între părinte (în special mamă) şi copil.

La capătul depărtat al poveştii iubirii, persoana iubită devine o parte importantă a neocortexului tău. După decenii petrecute împreună, în neocortex există un „celălalt" virtual, astfel încât putem să anticipăm pas cu pas ce va spune şi va face el. Modelele noastre neocorticale sunt pline de gânduri şi modele care reflectă cine este „celălalt". Când pierdem această persoană, literalmente pierdem o parte din noi înşine. Nu e doar o metaforă – toţi recognizatorii modelului de ansamblu, care sunt plini de modele ce reflectă persoana pe care o iubim, îşi schimbă, brusc, natura. Cu toate că pot fi considerate un mijloc nepreţuit de păstrare a celuilalt în viaţă înăuntrul nostru, vastele modele neocorticale ale unei persoane iubite pierdute devin brusc, din declanşatori ai încântării, declanşatori ai durerii.

Baza evolutivă a iubirii şi fazele ei nu reprezintă totul în lumea de astăzi. În mare măsură, am reuşit să eliberăm sexul de funcţia sa biologică, prin faptul că putem face copii fără sex şi, cu siguranţă, putem face sex fără să facem copii. Vasta majoritate a raporturilor sexuale se produce în scopuri senzuale şi relaţionale. Şi ne îndrăgostim, în mod curent, pentru alte motive decât cel de a face copii.

În mod similar, vasta cuprindere a expresiei artistice

de toate genurile care celebrează iubirea şi nenumăratele ei forme, datând încă din Antichitate, reprezintă de asemenea un scop de sine stătător. Şi exact capacitatea noastră de a crea aceste forme durabile de cunoaştere transcendentală – referitoare la iubire sau la orice altceva – determină specia noastră să fie unică.

Neocortexul reprezintă cea mai mare creaţie a biologiei. Poeziile de iubire – şi toate celelalte creaţii ale noastre – sunt cele mai mari creaţii ale neocortexului nostru.

Capitolul 7

NEOCORTEXUL DIGITAL INSPIRAT DE BIOLOGIE

Niciodată să nu te încrezi în ceva ce poate gândi independent dacă nu vezi unde îşi ţine creierul.

— Arthur Weasley, în J.K. Rowling, *Harry Potter şi prizonierul din Azkaban*

Nu, nu mă interesează crearea unui creier puternic. Urmăresc doar un creier mediocru, cam ca al preşedintelui Companiei Americane de Telefonie şi Telegraf.

— Alan Turing

Un computer ar merita să fie considerat inteligent dacă ar putea înşela un om făcându-l să creadă că e om.

— Alan Turing

Cred că, la sfârşitul secolului, folosirea cuvintelor şi opinia generală avizată se vor modifica atât de mult, încât vei putea vorbi despre maşini care gândesc fără să te temi că ai putea fi contrazis.

— Alan Turing

Şobolanul-mamă construieşte un cuib pentru puii săi chiar dacă nu a mai văzut un alt şobolan în viaţa ei.1 La fel, păianjenul îşi ţese pânza, coropişniţa îşi face propriul cocon, iar castorul construieşte un dig chiar dacă niciun contemporan nu le-a arătat vreodată cum se fac aceste lucruri. Asta nu înseamnă că aceste comportamente nu sunt învăţate, ci doar că aceste animale nu le-au învăţat în decursul unei singure vieţi, ci pe parcursul a mii de vieţi. Evoluţia comportamentului animalelor constituie pe bună

dreptate un proces de învăţare, dar este vorba de învăţarea la nivelul speciei, nu al individului, iar rodul acestui proces de învăţare este codificat în ADN.

Pentru a înţelege semnificaţia evoluţiei neocortexului, să ne gândim că evoluţia a accelerat foarte mult procesul de învăţare (cunoaştere Ierarhică) de la mii de ani la câteva luni (sau mai puţin). Chiar dacă milioane de animale dintr-o anumită specie de mamifere nu au reuşit să rezolve o problemă (necesitând o ierarhie de paşi), a fost nevoie ca doar unul singur să dea din întâmplare peste o soluţie. Pe urmă, această nouă metodă a fost copiată şi răspândită exponenţial în rândul populaţiei.

Ne aflăm acum în poziţia de a accelera din nou procesul de învăţare de mii sau de milioane de ori, prin trecerea de la inteligenţa biologică la inteligenţa nonbiologică. Din momentul în care un neocortex digital învaţă o abilitate, poate să transfere acel know-how în câteva minute sau chiar secunde. Printre nenumăratele exemple: la prima mea companie, Kurzweil Computer Products (actualmente Nuance Speech Technologies), pe care am fondat-o în anul 1973, am avut nevoie de ani întregi pentru a pregăti un set de computere de cercetare ca să recunoască literele tipărite dintr-un document, tehnologie numită omnifont (orice tip de font), recunoaştere optică a caracterelor (*Optical Character Recognition* – *OCR*). Această tehnologie se află într-un continuu proces de dezvoltare de aproape patruzeci de ani, actualul produs numindu-se *Omni Page de la Nuance*. Dacă vrei ca ordinatorul tău să recunoască litere tipărite, nu trebuie să pierzi ani de-a rândul pentru a-l învăţa acest lucru, aşa cum am făcut noi – poţi pur şi simplu să descarci modelele evoluate deja învăţate de computerele de cercetare sub formă de software. În anii 1980, am început programul de recunoaştere a vorbirii, iar acea tehnologie, care se află de asemenea într-un proces de continuă dezvoltare de câteva decenii, face parte din Siri. Şi din nou

poţi să descarci în câteva secunde modelele evoluate învăţate de computerele de cercetare pe parcursul mai multor ani.

În cele din urmă, vom crea un neocortex artificial care să dispună de întreaga anvergură şi flexibilitate a corespondentului lui uman. Gândeşte-te numai la beneficii! Circuitele electronice sunt de milioane de ori mai rapide decât circuitele noastre biologice. La început, va trebui să dedicăm această sporire a vitezei, pe de-a-ntregul, compensării relativei lipse de paralelism din computerele noastre, dar în cele din urmă neocortexul digital va fi mult mai rapid decât varianta biologică şi va continua să-şi sporească viteza.

Când ne vom îmbunătăţi propriul neocortex cu o versiune sintetică, nu va mai trebui să ne facem griji legate de cât de mult neocortex artificial suplimentar va avea loc fizic în corpul şi creierul nostru, pentru că, în cea mai mare parte a lui, el se va afla în nor, ca majoritatea mijloacelor de computare pe care le folosim astăzi. Am estimat anterior că, la nivelul neocortexului nostru biologic, avem ceva de ordinul a 300 de milioane de recognizatori de modele. Doar atâta s-a putut înghesui în cutia noastră craniană, chiar şi cu inovaţia evoluţionistă a unei frunţi înalte şi în condiţiile în care neocortexul ocupă aproximativ 80 la sută din spaţiul disponibil. Din momentul în care începem să ne gândim la nor, nu va mai exista vreo limită naturală – vom putea să folosim miliarde sau mii de miliarde de recognizatori, în esenţă cât de mulţi ne-ar trebui şi cât de mulţi ne poate asigura, în fiecare moment, Legea accelerării veniturilor.

Ca să înveţe o nouă abilitate, un neocortex digital are nevoie în continuare de multe repetări ale lecţiilor, la fel ca neocortexul biologic, dar, din momentul în care un singur neocortex digital din vreun colţ al lumii învaţă ceva, el poate să le împărtăşească această cunoaştere tuturor celorlalte neocortexuri digitale, fără întârziere. Fiecare dintre noi îşi va putea avea propriile extensii de neocortex

privat în nor, la fel cum astăzi beneficiem de propriile noastre stocuri de date personale.

În sfârşit, dar nu în cele din urmă, vom putea să salvăm o copie de rezervă a părţii digitale a inteligenţei noastre. După cum am văzut, nu e o simplă metaforă să afirmăm că în neocortexul nostru este stocată informaţie şi e înfiorător să ne gândim că nimic din această informaţie nu are salvată undeva o copie de rezervă. Desigur, există o cale prin care ne facem totuşi o asemenea copie de rezervă pentru o parte din informaţia din creierserіind-o pe hârtie. Capacitatea de a transfera cel puţin o parte din gândirea noastră pe un mediu care poate supravieţui trupurilor noastre biologice a reprezentat un uriaş pas înainte, dar un volum foarte mare de date din creierul nostru continuă să rămână vulnerabil.

Simulările creierului

O abordare a construirii unui creier digital constă în simularea precisă a unuia biologic. De exemplu, David Dalrymple (născut în 1991), doctorand în ştiinţele creierului la Harvard, intenţionează să simuleze creierul unui nematod (vierme rotund).2 Dalrymple a ales nematodul datorită sistemului său nervos relativ simplu, care însumează aproximativ 300 de neuroni şi pe care intenţionează să-l simuleze până la nivelul de detaliu al moleculelor. De asemenea, cercetătorul va realiza o simulare computerizată a corpului său şi a mediului, astfel încât nematodul lui virtual să poată vâna hrana (virtuală) şi să poată face restul lucrurilor la care sunt buni nematozii. Dalrymple spune că e posibil să fie primul upload complet de creier între un animal biologic şi unul virtual, care trăieşte într-o lume virtuală. La fel ca în cazul nematodului simulat, încă se mai discută dacă nematozii biologici sunt conştienţi, deşi, în lupta lor pentru hrănire, digestie, evitarea prădătorilor şi reproducere, au experienţe de care pot fi conştienţi.

La capătul opus al spectrului, Proiectul Blue Brain al lui Henry Markram intenţionează să simuleze creierul

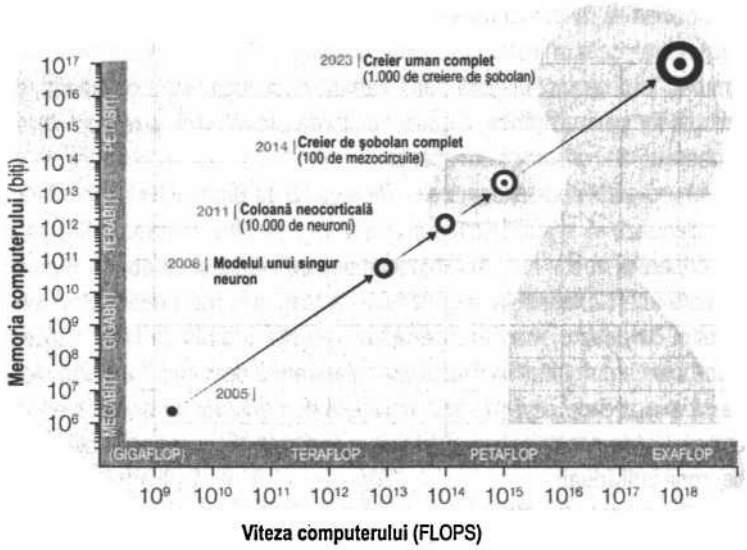
uman, inclusiv întregul neocortex şi regiuni ale creierului vechi, cum ar fi hipocampul, amigdala şi cerebelul. Simulările pe care intenţionează să le construiască Markram vor fi realizate la diferite niveluri de detaliu, mergând până la o simulare totală, la nivel molecular. După cum am anunţat în Capitolul 4, Markram a descoperit un modul-cheie de câteva zeci de neuroni care se repetă de nenumărate ori la nivelul neocortexului, demonstrând că învăţarea e realizată de aceste module şi nu de neuronii individuali.

Progresele lui Markram se acumulează într-un ritm exponenţial. A simulat un neuron în 2005, anul iniţierii proiectului său. În anul 2008, echipa lui a simulat întreaga coloană neocorticală a unui creier de şobolan, care consta în 10.000 de neuroni. Până în anul 2011, a ajuns la un total de 100 de coloane, totalizând un milion de neuroni, pe care îi numeşte mezocircuit. O controversă privitoare la cercetările lui Markram ridică întrebarea cum se poate verifica dacă simulările sunt precise. Pentru a răspunde, simulările vor trebui să demonstreze că învăţă, aşa cum voi arăta în cele ce urmează.

Markram intenţionează să simuleze un creier de şobolan complet de 100 de mezocircuite, totalizând 100 de milioane de neuroni şi în jur de trei miliarde de sinapse, până în anul 2014. În cursul unei discuţii în cadrul conferinţei TED din 2009 de la Oxford, Markram a spus: „Nu este imposibil de construit un creier uman şi o putem fac peste zece ani”.3 Cea mai recentă ţintă a lui în direcţia simulării totale a creierului uman o reprezintă anul 2023/

Markram şi echipa lui îşi întemeiază modelul pe analize detaliate, anatomice şi electrochimice, ale unor neuroni reali. Folosind un dispozitiv automat, pe care l-au denumit *robot patch-clamp*, ei măsoară canalele specifice de ioni, neurotransmiţători şi enzime care sunt răspunzătoare pentru activitatea electrochimică de la nivelul fiecărui neuron. Potrivit lui Markram, sistemul lor automat a reuşit să realizeze în şase luni echivalentul a

treizeci de ani de analize. În cursul acestor analize, au fost observate unităţile de genul „memorie Lego”, adică unităţile funcţionale de bază ale neocortexului.

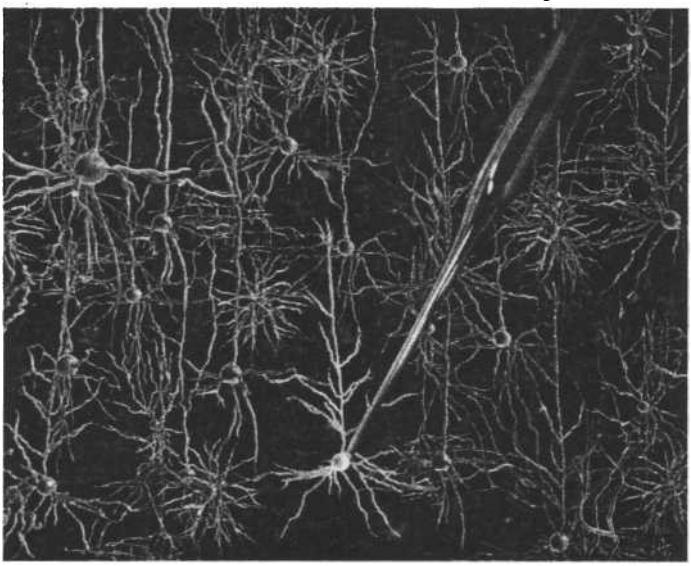


Progresele actuale şi anticipate ale Proiectului Blue Brain de simulare a creierului

Contribuţii semnificative la tehnologia tehnicii robotizate au fost înregistrate de specialistul în neuroştiinţă de la MIT Ed Boyden, de profesorul de inginerie mecanică de la Georgia Tech Craig Forest şi de un absolvent al lui Forest, Suhasa Kodandaramaiah. Ei au efectuat demonstraţii cu ajutorul unui sistem automat având precizia de un micrometru. Acesta poate să realizeze scanarea ţesutului neuronal la distanţe foarte mici fără să lezeze membranele delicate ale neuronilor. „E un lucru pe care un robot îl poate face, dar omul, nu”, a comentat Boyden.

Ca să revenim la Markram, după ce acesta a simulat coloana neocorticală, a fost citat ca spunând: „Acum mai trebuie doar să mărim la scară”.5 Mărirea la scară este în mod cert un factor important, dar mai există o problemă-cheie – şi anume învăţarea. Dacă se doreşte ca creierul produs de Proiectul Blue Brain „să vorbească, să aibă inteligenţă şi să se comporte foarte asemănător unui om”,

aşa cum şi-a descris Markram obiectivul în cursul unui interviu acordat BBC în anul 2009, atunci va trebui ca acesta să aibă în neocortexul său simulat un conţinut suficient de mare pentru realizarea tuturor acestor funcţii.6 După cum poate atesta oricine a încercat să poarte o conversaţie cu un nou-născut, înainte ca acest lucru să devină fezabil, trebuie să înveţe foarte mult.



Vârful robotului *patchdamping* pus la punct la MIT şi Georgia Tech, în timp ce scanează ţesutul neuronal

Există două mijloace evidente prin care se poate atinge acest obiectiv în cazul unui creier simulat asemănător lui Blue Brain. Unul ar fi ca respectivul creier să înveţe acest conţinut aşa cum o face un creier de om. La început, poate fi asemenea unui nou-născut cu o capacitate de acumulare a cunoaşterii ierarhice înnăscută şi cu anumite transformări preprogramate în regiunile senzoriale preprocesatoare. Însă învăţarea ce are loc între un copil uman şi un om care poate întreţine o conversaţie ar trebui să se desfăşoare într-o manieră comparabilă şi în cazul învăţării nonbiologice. Problema în cazul acestei abordări este că un creier simulat la nivelul de detaliu anticipat pentru Blue Brain e de aşteptat să funcţioneze în

timp real cel mai curând la începutul anilor ’2020. Chiar şi funcţionarea în timp real poate fi prea lentă dacă cercetătorii nu sunt pregătiţi să aştepte un deceniu sau două pentru ca produsul lor să ajungă la paritate intelectuală cu un adult uman, deşi performanţa în timp real va fi permanent mai bună, pe măsură ce computerele vor continua să evolueze sub aspectul raportului preţ/performanţă.

Cealaltă abordare presupune acţiunea de a se lua mai multe creiere umane biologice care au achiziţionat deja suficientă cunoaştere pentru a conversa într-un limbaj inteligibil, pentru a se comporta în general de o manieră matură şi pentru a-şi copia modelele neocorticale în creierul simulat, în cazul acesta, problema e că metoda necesită o tehnologie de scanare non-invazivă şi nondistructivă cu rezoluţie spaţială şi temporală suficientă, dotată cu viteză pentru a îndeplini o asemenea sarcină rapid şi complet Nu m-aş aştepta ca o asemenea tehnologie de *upload* să fie disponibilă până în jurul anilor ’2040. (Cerinţele computaţionale pentru simularea creierului cu o asemenea precizie, pe care o estimez la 1.019 calcule pe secundă, va fi accesibilă unui supercomputer, potrivit proiecţiilor mele, înspre începutul anilor ’2020; însă tehnologiile nondistructive de scanare a creierului vor avea nevoie de mai mult timp.)

Există şi o a treia abordare, la care, cred eu, ar putea să recurgă proiectele de simulare a creierului de tipul Blue Brain. Modelele moleculare se pot simplifica prin crearea unor echivalente funcţionale la diferite niveluri de specificitate, mergând de la metoda mea funcţională algoritmică (pe care o descriu în această carte) până la simulări mai apropiate de simulările moleculare totale. Viteza de învăţare poate fi, prin această tehnică, accelerată cu un ordin de mărime de sute sau mii, în funcţie de gradul de simplificare folosit. Pentru creierul simulat, poate fi conceput un program educaţional (cu ajutorul modelului funcţional) pe care acesta îl poate învăţa relativ rapid. Pe

urmă, întreaga simulare moleculară poate fi substituită cu modelul simplificat, care îşi foloseşte în continuare cunoştinţele acumulate. Ulterior, putem simula învăţarea cu modelul molecular complet, la o viteză mult mai redusă.

Specialistul american în ştiinţa computerelor Dharmendra Modha şi colegii săi de la IBM au creat o simulare celulă cu celulă a unei porţiuni din neocortexul vizual uman cuprinzând 1, 6 miliarde de neuroni virtuali şi 9 mii de miliarde de sinapse, echivalentul unui neocortex de pisică. Această simulare funcţionează de 100 de ori mai lent pe un supercomputer IBM Blue Gene/P, format din 147.456 de procesoare. Cercetarea a primit Premiul Gordon Bell din partea Asociaţiei pentru Maşini de Computare.

Scopul unui proiect de simulare a creierului precum Blue Brain şi al unui neocortex simulat ca al lui Modha este în mod expres acela de a rafina şi confirma un model funcţional. Inteligenţa artificială la nivel uman va folosi, în principal, modelul funcţional algoritmic discutat în această carte. Pe de altă parte, simulările moleculare ne vor ajuta să perfecţionăm acest model şi să înţelegem pe deplin ce detalii sunt importante. În cursul cercetărilor noastre privitoare la punerea la punct a tehnologiei de recunoaştere a vorbirii din anii 1980 şi 1990, am reuşit să ne rafinăm algoritmii din momentul în care am înţeles transformările propriu-zise realizate de nervul auditiv şi de porţiunile timpurii ale cortexului auditiv. Deşi modelul nostru funcţional a fost perfect, abia atunci când vom înţelege cu precizie cum este el implementat practic în creierul nostru biologic, ni se vor dezvălui cunoştinţele importante referitoare la funcţionalitatea şi disfuncţionalitatea umană.

Vom avea nevoie de date detaliate referitoare la creierul real pentru a realiza simulări bazate pe biologie. Echipa lui Markram îşi adună propriile date. Există proiecte pe scară largă de colectare a acestor date şi de punere a lor la dispoziţia oamenilor de ştiinţă. De exemplu,

Laboratorul Cold Spring Harbor din New York a strâns, în urma scanării unui creier de mamifer (şoarece), 500 de terabiţi de informaţii, pe care le-a făcut accesibile în luna iunie 2012. Proiectul îi permite utilizatorului să exploreze un creier tot aşa cum Google Earth permite explorarea suprafeţei planetei. Te poţi deplasa în jurul întregului creier mărindu-l pentru a vedea neuronii individuali şi conexiunile lor. Poţi scoate în evidenţă o singură conexiune şi pe urmă îi poţi urmări traseul prin creier.

Şaisprezece secţii ale Institutelor Naţionale pentru Sănătate s-au reunit şi au sponsorizat o iniţiativă majoră, intitulată Human Genome Project – Proiectul Genomului Uman –, cu suma de 38, 5 milioane de dolari.7 Derulat de Universitatea Washington din St. Louis, Universitatea din Minnesota, Universitatea Harvard, Spitalul General Massachusetts şi Universitatea California din Los Angeles, proiectul îşi propune să creeze o hartă similară tridimensională a conexiunilor din creierul uman. El foloseşte o varietate de tehnologii de scanare non-invazive, inclusiv noi forme de RMN, magnetoencefalografie (măsurarea câmpurilor magnetice produse de activitatea electrică a creierului) şi tractografie prin difuziune (metodă de trasare a căilor fasciculelor de fibre din creier). După cum subliniez în Capitolul 10, rezoluţia spaţială a scanării non-invazive a creierului se îmbunătăţeşte cu o rată exponenţială. Cercetările lui Van J. Wedeen şi ale colegilor săi de la Spitalul General Massachusetts, care au relevat o structură în grilă extrem de regulată a cablării neocortexului (vezi Capitolul 4), reprezintă un rezultat timpuriu al acestui proiect.

Specialistul în neuroştiinţă computaţională Anders Sandberg (născut în anul 1972) de la Universitatea Oxford şi filosoful suedez Nick Bostrom (născut în anul 1973) au scris o lucrare cuprinzătoare intitulată *Whole Brain Emulation*: *A Roadmap – Imitarea creierului în ansamblu*: *îndrumar*, care detaliază

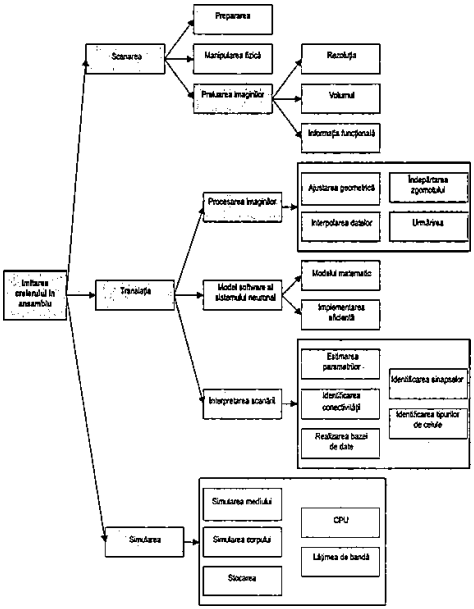
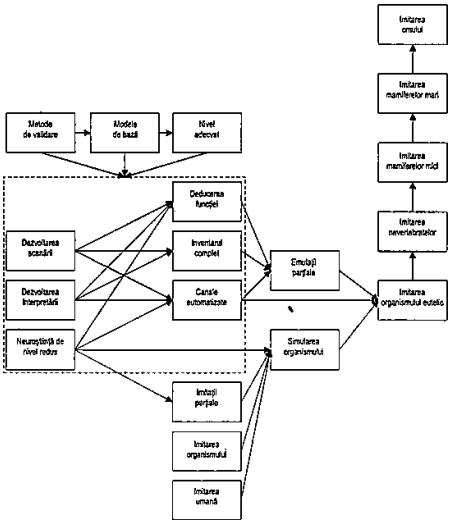


Diagrama capacităţilor tehnologice necesare pentru imitarea creierului în ansamblul său, din *Whole Brain Emulation*: *A Roadmap – Imitarea creienilui în ansamblu*: *indnimar*, de Anders Sandberg şi Nick Bostrom cerinţele referitoare la imitarea creierului uman (şi a altor tipuri de creiere) pe diferite niveluri de specificitate – de la modelele funcţionale de nivel superior, la simularea moleculelor.8 Comunicarea nu oferă o cronologie, însă descrie cerinţele necesare pentru simularea diferitelor tipuri de creier la diferite niveluri de precizie sub aspectul scanării cerebrale, realizării modelelor, stocării şi computării. Comunicarea proiectează evoluţii exponenţiale continue în toate aceste domenii ale capabilităţii şi susţine că exigenţele privind simularea creierului uman la un nivel superior de detaliu sunt adecvate.



Rezumatul cărţii *Whole Brain Emulation*: *A Roadmap – Imitarea creierului în ansamblu*: *îndrumar*, de Anders Sandberg şi Nick Bostrom

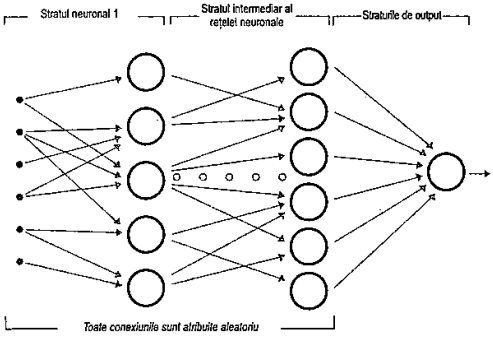
128 RAY KURZWEIL

Reţele neuronale în anul 1964, la vârsta de şaisprezece ani, i-am scris lui Frank Rosenblatt (1928 – 1971), profesor la Universitatea Cornell, adresându-i întrebări cu privire la o maşină denumită Mark 1 Perception. O crease cu patru ani înainte şi era descrisă ca având proprietăţi asemănătoare creierului. El m-a invitat să-l vizitez şi să încerc maşina.

Percepţional era construit din ceea ce Rosenblatt pretindea că erau mqdele electronice ale neuronilor. Imputul consta în valori aranjate în două dimensiuni. Pentru vorbire, o dimensiune reprezenta frecvenţa iar cealaltă timpul, astfel încât fiecare valoare ilustra intensitatea unei frecvenţe într-un anumit moment. Pentru imagini, fiecare punct era un pixel dintr-o imagine bidimensională. Fiece punct de un anumit imput era conectat aleatoriu la inputurile primului strat de neuroni simulaţi. Fiecare conexiune avea asociată o tărie sinaptică pe măsura importanţei ei, iar aceasta era stabilită iniţial la o valoare aleatorie. Fiece neuron aduna semnalele care ajungeau la el. În caz că semnalul combinat depăşea un

anumit prag, neuronul emitea un, semnal la conexiunea sa de output; dacă semnalul combinat nu depăşea pragul, neuronul nu emitea semnalul, aşa că outputul lui era zero. Outputul fiecărui neuron era conectat aleatoriu la inputurile neuronilor din următorul strat. Mark 1 Perception avea trei straturi, care puteau fi organizate într-o largă varietate de configuraţii. De exemplu, un strat putea să transmită feedback la alt strat anterior. În stratul superior, outputul unuia sau al mai multor neuroni, de asemenea aleşi aleatoriu, asigura răspunsul. (Pentru o descriere algoritmică a reţelelor neuronale, vezi această notă de final.) 9

Deoarece deconectarea reţelei neuronale şi greutăţile sinaptice sunt stabilite la început aleatoriu, răspunsurile unei asemenea reţele fără pregătire vor fi şi ele aleatorii. Prin urmare, reţeaua neuronală trebuie să-şi înveţe materia, exact cum o face creierul mamiferelor, al cărui model se presupune că l-ar reprezenta. Reţeaua neuronală debutează într-o stare de ignoranţă; profesorul – care poate fi un om, un program de computer sau o reţea neuronală mai matură, care şi-a învăţat deja lecţiile – recompensează reţeaua-student atunci când aceasta generează outputul corect şi o pedepseşte când nu o face. Acest feedback este folosit la rândul lui de către reţeaua neuronală student pentru ajustarea tăriei fiecărei conexiuni


Stratul neuronal 1
Stratul intermediar al rețelei neuronale
Straturile de output
Toate conexiunile sunt atribuite aleatoriu

Interneuronale. Conexiunile care concordă cu răspunsul corect sunt întărite. Cele care dau răspunsuri

greşite sunt slăbite.

Cu timpul, reţeaua neuronală se organizează pentru a da răspunsurile corecte fără pregătire. Experimentele au arătat că reţelele neuronale îşi pot învăţa materia chiar şi cu profesori care nu prezintă garanţii. Chiar dacă dascălul are dreptate în numai 60 la sută dintre cazuri, reţeaua neuronală student îşi învaţă lecţiile cu o acurateţe ce se apropie de 100 la sută.

Pe de altă parte însă, au devenit evidente limitările în privinţa materialului pe care Percepţional era capabil să-l înveţe repede. Când l-am vizitat pe profesorul Rosenblatt, în anul 1964, am încercat, modificări simple ale inputului. Sistemul era setat să recunoască litere tipărite şi putea să le recunoască precis. Se descurca destul de bine la autoasociere (respectiv, putea să recunoască literele chiar dacă erau acoperite părţi din ele), dar se descurca mai puţin bine cu invarianţa (respectiv generalizarea indiferent de mărime şi font, care îi provocau confuzii).

În cursul ultimei jumătăţi a anilor 1960, aceste reţele neuronale au devenit enorm de populare, iar domeniul „conexionismului" a pus stăpânire pe mai mult de jumătate din domeniul inteligenţei artificiale. Între timp, o abordare mai tradiţională a inteligenţei artificiale a inclus tentative de programare a soluţiilor pentru probleme specifice, cum ar fi recunoaşterea proprietăţilor invariabile ale literelor tipărite.

O altă persoană pe care am vizitat-o în anul 1964 a fost Mârvin Minsky (născut în anul 1927), unul dintre fondatorii domeniului inteligenţei artificiale, în ciuda faptului că, în anii 1950, făcuse personal o serie de cercetări de pionierat în domeniul reţelelor neuronale, era îngrijorat de marele val de Interes faţă de această tehnică. Parte din farmecul reţelelor neuronale consta în faptul că se presupunea că acestea nu necesitau programare – că puteau să înveţe soluţiile la probleme pe cont propriu. În anul 1965 am Intrat la MIT, avându-l pe profesorul Minsky drept mentor, şi i-am împărtăşit scepticismul în legătură cu

nebunia „conexionismului”.

În anul 1969, Minsky şi Seymour Papert (născut în anul 1928), cofondatori ai Laboratorului de Inteligenţă Artificială de la MIT, au scris o carte intitulată *Percetrons-Perceptroni*, care prezenta o singură teoremă centrală şi anume că Percepţional era în mod inerent incapabil să determine faptul că o imagine era conectată sau nu. Cartea a provocat un uragan. Determinarea faptului că o imagine este conectată sau ba reprezintă o sarcină pe care oamenii o pot îndeplini cu uşurinţă; este, de asemenea, un proces simplu să programezi un computer pentru a face această diferenţiere. Faptul că Percepţional nu putea să realizeze acest lucru a fost considerat de mulţi un neajuns fatal.

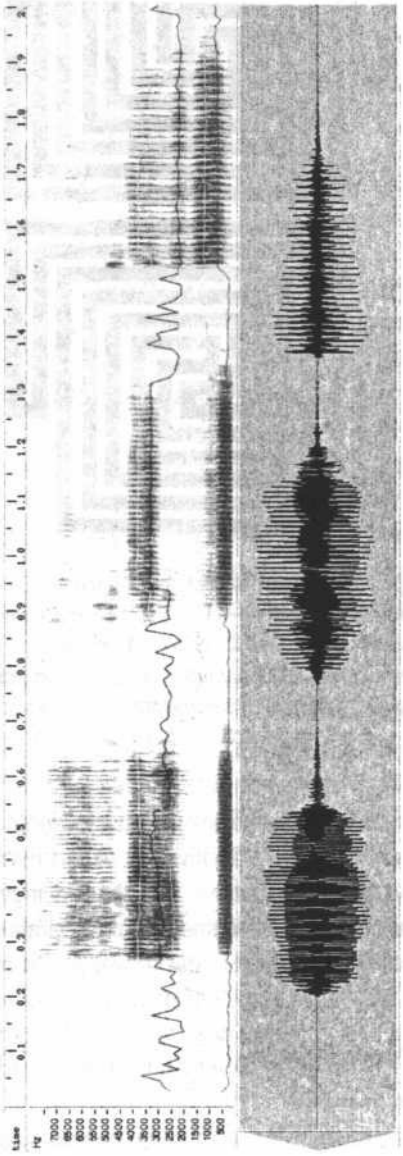
Oricum, *Percepţions-Perceptroni* a fost interpretată pe scară largă ca Implicând mai multe decât o făcea în realitate. Teorema lui Minsky şi Papert se aplică doar unui tip particular de reţea neuronală, denumit feedforward (categorie care include şi Percepţional lui Rosenblatt); alte tipuri de reţele neuronale nu aveau această limitare. Totuşi, cartea a reuşit să ucidă în mare măsură finanţarea cercetărilor pe tema reţelelor neuronale în cursul anilor 1970. Domeniul a revenit însă în anii 1980 cu tentative de folosire a unor modele mai realiste ale neuronilor biologici şi a unora care evitau limitările pe care le implica Teorema Perception a lui Minsky-Papert. Cu toate acestea, abilitatea neocortexului de a rezolva problema invarianţei, cheia forţei sale, a rămas în continuare insolvabilă pentru conexioniştii renăscuţi ai domeniului.

Codificarea sporadică: *cuantificarea vectorilor*

La începutul anilor 1980, am demarat un proiect dedicat unei alte probleme clasice de recunoaştere a modelelor: înţelegerea vorbirii umane. La

Două imagini de pe coperta cărţii *Perceptions* – *Percepţioni* a lui Marvin Minsky şi Seymour Papert. Imaginea de sus nu este conectată (respectiv, zonele întunecate constau din două părţi deconectate). Imaginea de jos e conectată. Omul poate determina cu uşurinţă acest lucru, la fel ca un program software simplu. Un Perception cu feedforward, cum era Percepţional Mark 1 al lui Frank Rosenblatt, nu putea să facă această determinare început, am folosit abordări tradiţionale pentru inteligenţa artificială prin programarea directă a cunoştinţelor la nivel de expert referitoare la unităţile fundamentale de vorbire – fonemele – şi la regulile lingvistice privitoare la modul în care oamenii înşiruiesc fonemele pentru a forma cuvinte şi expresii. Fiecare fonem are modele distincte de frecvenţe. De exemplu, ştiam că vocale precum „e” şi „a” sunt caracterizate de anumite frecvenţe rezonante denumite formatori, cu o proporţie caracteristică de formatori pentru fiecare fonem. Sunetele sibilante „z” şi „s” sunt definite de o explozie sonoră care cuprinde mai multe frecvenţe.

Am captat vorbirea sub formă de undă, pe care am convertit-o în mai multe benzi de frecvenţă (percepute ca înălţime), folosind o bancă de filtre de frecvenţe. Rezultatul acestei transformări putea fi vizualizat şi a fost denumit spectrogramă.

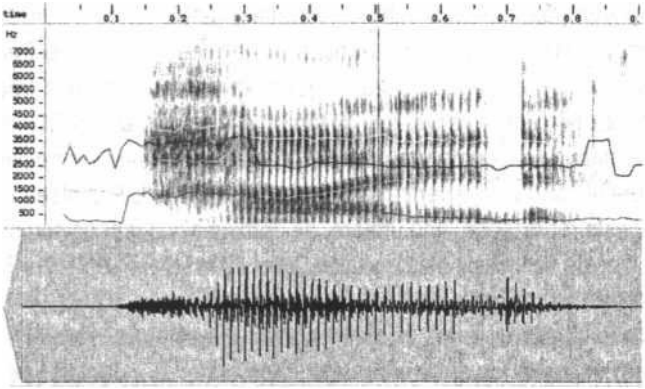


Spectrograma pentru trei vocale. De la stânga la

dreapta: ţi] ca în", [u] ca în i [al ca în „*ah*".

Axa y reprezintă frecvenţa sunetului. Cu cât banda e mai întunecată, cu atât energia acustică pe acea frecvenţă este mai mare

Banca de filtre copiază ceea ce face cohleea umană, acesta fiind pasul iniţial în procesarea biologică a sunetului. Programul software a identificat la început fonemele pe baza distincţiei dintre modelele de frecvenţă şi pe urmă a recunoscut cuvinte pe baza identificării succesiunilor caracteristice de foneme.



Spectrogramă a unei persoane care rosteşte cuvântul „hide". Liniile orizontale arată formatorii, care sunt frecvenţe susţinute, cu o energie deosebit de mare10

Rezultatul a fost parţial încununat de succes. Puteam să ne învăţăm dispozitivul să înveţe modele pentru o anumită persoană care folosea un vocabular de dimensiuni moderate, măsurat în câteva mii de cuvinte. Când am încercat să recunoaştem zeci de mii de cuvinte, să lucrăm cu mai mulţi vorbitori şi să permitem vorbirea continuă (respectiv vorbirea fără pauze între cuvinte), am dat peste problema invarianţei. Diferiţi oameni pronunţă aceleaşi foneme diferit – de exemplu, fonemul „e" al unei persoane poate să semene cu fonemul „ah" al alteia. Chiar şi una şi aceeaşi persoană prezintă inconsecvenţă în modul de rostire a unui anumit fonem. Modelul unui fonem a fost adeseori afectat de alte foneme învecinate. Multe foneme au fost eliminate complet. Pronunţia cuvintelor (respectiv

felul în care sunt înşirate fonemele pentru formarea cuvintelor) era de asemenea extrem de variabilă şi dependentă de context. Regulile lingvistice pe care le programasem se prăbuşeau şi nu puteau ţine pasul cu extrema variabilitate a limbajului vorbit.

La vremea respectivă, mi-a devenit clar că esenţa modelului uman şi a recogniţiei conceptuale se bazează pe ierarhii. Acest fapt este cu certitudine evident pentru limbajul uman, care constituie o ierarhie elaborată de structuri. Dar care este elementul de la baza structurilor? Aceasta a fost prima întrebare pe care am avut-o în vedere atunci când am căutat mijloacele de recunoaştere automată a vorbirii umane integrale.

Sunetul pătrunde în ureche ca vibraţie a aerului şi este convertit, de aproximativ 3.000 de celule cu perişori din cohlee, în benzi de frecvenţă multiple. Fiecare celulă cu perişori e acordată pentru o anumită frecvenţă (observă că percepem frecvenţele ca tonuri) şi funcţionează ca filtru de frecvenţă, emiţând un semnal ori de câte ori există un sunet cu frecvenţa ei de rezonanţă sau cu frecvenţă apropiată. Când părăseşte cohleea umană, sunetul este reprezentat, prin urmare, de aproximativ 3.000 de semnale separate, fiecare semnificând intensitatea variabilă în timp a unei benzi înguste de frecvenţă (cu o suprapunere substanţială între benzi).

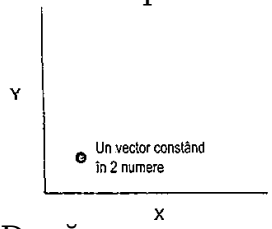
Cu toate că devenise evident că creierul era masiv paralel, mie mi s-a părut imposibil ca el să efectueze împerecherea modelelor pe 3.000 de semnale auditive separate. M-am îndoit că evoluţia a putut fi într-atât de eficientă. Actualmente ştim că, într-adevăr, se realizează o reducere substanţială a datelor la nivelul nervului auditiv înainte ca semnalele să ajungă la cortex.

La recognizatorii noştri de vorbire bazaţi pe software, am folosit de asemenea filtre implementate ca software – şaisprezece, ca să fiu mai exact (număr pe care ulterior l-am mărit la treizeci şi doi, deoarece am constatat că nu exista niciun beneficiu dacă depăşeam acest număr).

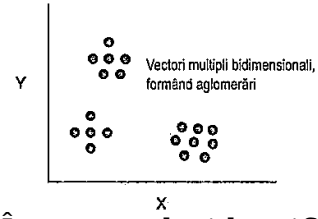
Prin urmare, în sistemul nostru, fiecare punct din timp era reprezentat de şaisprezece numere. Am fost obligaţi să reducem aceste şaisprezece fluxuri de date la unul singur şi în acelaşi timp să accentuăm caracteristicile semnificative în recunoaşterea vorbirii.

Noi am folosit o tehnică matematică optimă, numită cuantificarea vectorială. Să spunem că, în orice moment de timp, sunetul (cel puţin de la o ureche) a fost reprezentat de software-ul nostru prin şaisprezece numere diferite: respectiv outputul celor şaisprezece filtre de frecvenţă. (La nivelul sistemului auditiv uman, totalul ar fi de aproximativ 3.000, cifră reprezentând outputul celor 3.000 de celule interioare cohleare cu perişori.) în terminologie matematică, fiecare asemenea set de numere (indiferent că ar fi 3.000 în cazul biologic sau 16 în implementarea noastră software) se numeşte *vector*.

Pentru simplificare, să luăm procesul cuantificării cu vectori pentru două numere. Fiecare vector poate fi considerat un punct la nivel bidimensional.

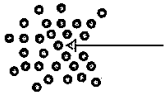


Dacă avem o mostră foarte mare de asemenea vectori şi îi plotăm, este posibil să observăm formarea de aglomerări



În scopul identificării aglomerărilor, trebuie să decidem cât de multe vom accepta. În cadrul proiectului nostru, am acceptat în general 1.024 de aglomerări, astfel încât să le putem număra şi să putem atribui fiecărei

aglomerări o etichetă de 10 biţi (pentru că 210 = 1.024). Mostra noastră de vectori ilustrează diversitatea pe care o aşteptam. Am hotărât cu titlu de încercare ca primii 1.024 de vectori să fie aglomerări de puncte individuale. Pe urmă am luat vectorul cu numărul 1.025 şi am găsit punctul de care era cel mai aproape. Dacă această distanţă este mai mare decât distanţa cea mai mică dintre orice pereche de 1.024 puncte, considerăm că ea este punctul de început al unei noi aglomerări. Pe urmă, reducem cele mai apropiate două aglomerări (de puncte individuale) la o singură aglomerare. În acest fel, tot ne-au mai rămas 1.024 de aglomerări. După procesarea vectorului cu numărul 1.025, una dintre aceste aglomerări are de-acum mai mult de un singur punct. Continuăm să procesăm punctele în acest fel, păstrând în permanenţă 1.024 de aglomerări. După ce am procesat toate punctele, reprezentăm fiecare aglomerare multipunct prin centrul geometric al punctelor din aglomerarea respectivă.



O aglomerare de puncte. O reprezentăm folosind un singur punct, care este centrul geometric al tuturor punctelor aglomerării

Continuăm acest proces iterativ până când folosim toate punctele mostrei. De obicei, procesăm un milion de puncte în aglomerări de câte 1.024 (210); am mai folosit şi aglomerări de 2.048 (211) sau 4.096 (212). Fiecare aglomerare este reprezentată de un vector situat în centrul geometric al tuturor punctelor din aglomerarea respectivă. Astfel, totalul distanţelor tuturor punctelor din aglomerare faţă de punctul central al aglomerării este cel mai mic posibil.

Rezultatul acestei tehnici este că, în loc să existe un milion de puncte de la care am început (şi un număr mai mare chiar de posibile puncte), am redus datele la numai 1.024 de puncte, care folosesc în mod optim spaţiul posibilităţilor. Părţilor aparţinând acestui spaţiu care nu sunt folosite niciodată nu li se atribuie vreo aglomerare.

Pe urmă, atribuim un număr fiecărei aglomerări (în cazul nostru, de la 0 la 1.023). Acest număr este

reprezentarea redusă „cuantificată” a acelei aglomerări, motiv pentru care tehnica se numeşte cuantificare vectorială. Orice nou imput care va sosi în viitor va fi reprezentat prin numărul aglomerării al cărei centru este cel mai apropiat de acest nou vector de imput.

Acum, putem precomputa un tabel cu distanţele de la punctele centrale ale fiecărei aglomerări până la fiecare dintre celelalte puncte centrale. În acest mod, avem instantaneu la dispoziţie distanţa noului vector de imput (pe care îl reprezentăm prin acest punct cuantificat – cu alte cuvinte, prin numărul aglomerării de al cărei centru este cel mai apropiat punctul) pentru fiecare aglomerare. Deoarece reprezentăm punctele doar prin aglomerările cele mai apropiate de ele, cunoaştem de-acum distanţa de la acest punct până la orice alt punct care ar putea să apară.

Am descris tehnica de mai înainte folosind vectori cu numai două numere fiecare, dar este absolut analog cu exemplul mai simplu dacă se lucrează cu vectori cu şaisprezece elemente. Pentru că am ales vectori cu şaisprezece numere reprezentând şaisprezece benzi diferite, fiecare punct din sistemul nostru are un punct în spaţiul cu şaisprezece dimensiuni. Nouă ne este greu să ne imaginăm un spaţiu cu mai mult de trei dimensiuni (poate patru, dacă includem şi timpul), dar matematica nu are asemenea inhibiţii.

Prin intermediul acestui proces, am realizat patru lucruri. Primul, am redus foarte mult complexitatea datelor. Al doilea, am redus datele în şaisprezece dimensiuni la date unidimensionale (respectiv, fiecare mostră este acum un singur număr). Al treilea, ne-am îmbunătăţit capacitatea de a găsi caracteristici invariabile, deoarece scoatem în evidenţă porţiuni ale spaţiului posibilelor sunete care transmit cele mai multe informaţii. Cea mai mare parte a combinaţiilor de frecvenţe e fizic imposibilă sau foarte puţin probabilă, astfel încât nu există niciun motiv pentru care să se aloce unor combinaţii

improbabile de inputuri un spaţiu egal cu acela pentru combin naţiile probabile. Această tehnică reduce datele la posibilităţi cu aceeaşi probabilitate. Al patrulea beneficiu este acela că putem folosi recognizatori de modele unidimensionali, chiar dacă datele originale constau în mult mai multe dimensiuni. Aceasta s-a dovedit a fi abordarea cea mai eficientă pentru utilizarea resurselor disponibile de computaţie.

Citirea gândurilor cu Modele Markov Ascunse

Prin cuantificarea vectorială, am simplificat datele într-un mod ce pune accentul pe caracteristicile-cheie, însă tot ne mai trebuia un mod de reprezentare a ierarhiei caracteristicilor invariabile care să facă logică noua informaţie. Pentru că în acel moment lucrasem deja în domeniul recognizatorilor de modele timp de douăzeci de ani, ştiam că reprezentările unidimensionale erau mult mai puternice, eficiente şi pretabile la rezultate invariabile. Neocortexul nu era prea bine cunoscut la începutul anilor 1980, dar, pe baza experienţei pe care o aveam cu o multitudine de probleme de recunoaştere, am presupus că era posibil ca şi creierul să reducă datele multidimensionale (provenite fie de la ochi, fie de la urechi sau piele) folosind o reprezentare unidimensională, mai ales că se dezvoltau conceptele privitoare la ierarhizarea neocortexului.

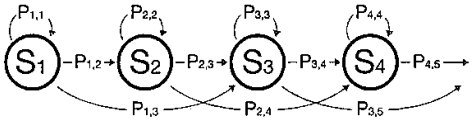
Pentru problema recunoaşterii vorbirii, organizarea informaţiei din semnalul vorbit a părut să însemne o ierarhie de modele, fiecare model fiind reprezentat de un şir liniar de elemente, cu o direcţie de înaintare. Fiece element al modelului putea fi un alt model de la un nivel inferior sau o unitate fundamentală de imput (care, în cazul recunoaşterii vorbirii, ar fi vectorii noştri cuantificaţi).

Îţi vei da seama, cititorule, că această situaţie este în concordanţă cu modelul neocortexului pe care ţi l-am prezentat mai înainte. Vorbirea umană, aşadar, este produsă de o ierarhie de modele liniare din creier. Dacă

am putea să examinăm pur şi simplu aceste modele din creierul persoanei care vorbeşte, totul s-ar rezuma la o simplă problemă de împerechere a noilor emisii verbale cu modelele cerebrale şi s-ar putea înţelege ce vrea să spună persoana respectivă. Din păcate, nu avem acces direct la creierul vorbitorului – singura informaţie pe care o obţinem fiind ceea ce spune el. Desigur, exact acesta este şi scopul limbajului vorbit-vorbitorul îşi împărtăşeşte nişte idei prin vorbire.

Şi atunci m-am întrebat: există vreo tehnică matematică prin care e posibil să deducem modelele din creierul vorbitorului pe baza cuvintelor rostite? Desigur, o singură repriză vorbită nu ar fi suficientă, dar, dacă am avea un număr mare de mostre, am putea să folosim această informaţie pentru a citi practic modelele din neocortexul vorbitorului (sau cel puţin să formulăm ceva echivalent matematic care să ne permită să recunoaştem noile reprize de vorbire)?

Oamenii nu reuşesc să îşi dea seama cât de minunată poate fi matematica – reţine, cititorule, că posibilitatea noastră de a căuta mare parte din cunoaşterea umană într-o fracţiune de secundă folosind motoarele de căutare se bazează pe o tehnică matematică. În cazul problemei, legate de recunoaşterea vorbirii, cu care mă confruntam la începutul anilor 1980, s-a dovedit că Modelele Ascunse ale lui Markov răspundeau cât se poate de bine cerinţelor. Matematicianul rus Andrei Andreeevici Markov (1856 – 1922) a construit o teorie matematică a succesiunilor ierarhice de stări. Modelul se baza pe posibilitatea parcurgerii stărilor într-un singur şir, iar, dacă acest lucru se realiza cu succes, pe generarea unei stări pe următorul nivel superior al ierarhiei. Sună familiar?



Un exemplu simplu de strat al unui Model Markov Ascuns. Cu S, până la S_4 am notat stările interne.

ascunse”. Tranziţiile P, (reprezintă, fiecare, probabilitatea de trecere de la starea S (la starea S. Aceste probabilităţi sunt determinate de sistemul care învaţă din datele de pregătire (inclusiv în cursul utilizării propriu-zise). O nouă succesiune (ca de pildă o nouă repriză de vorbire) este comparată cu aceste probabilităţi pentru a se determina probabilitatea ca modelul să producă succesiunea

Modelul lui Markov include probabilităţi de apariţie cu succes a fiecăreia dintre aceste stări. El a mers şi mai departe şi a formulat ipoteza unei situaţii în care un sistem are o asemenea ierarhie de succesiuni liniare de stări, dar ele nu pot fi examinate direct – de unde şi denumirea de Modele Markov Ascunse. Nivelul cel mai de jos ai ierarhiei emite semnale, acestea reprezentând tot ce ne este permis să vedem. Markov a furnizat o tehnică matematică de computare a probabilităţilor pentru fiecare tranziţie pe baza output-ului observat. Metoda a fost ulterior rafinată de Norbert Wiener, în anul 1923. Perfecţionarea lui Wiener a asigurat, de asemenea, un mijloc de determinare a conexiunilor pentru Modelul Markov; în esenţă, orice conexiune cu o probabilitate prea redusă era considerată inexistentă. În esenţă, acesta este modul în care taie neocortexul uman conexiunile – dacă sunt rar sau niciodată folosite, sunt considerate improbabile şi sunt retezate. În cazul nostru, outputul observat este semnalul verbal creat de persoana care vorbeşte, iar probabilităţile de stare şi conexiunile Modelului Markov constituie ierarhia neocorticală care l-a produs.

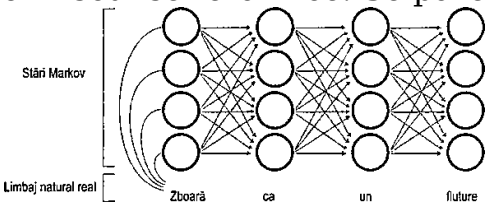
Eu am imaginat un sistem prin intermediul căruia să luăm mostre verbale umane, să le aplicăm tehnica Modelului Markov Ascuns pentru determinarea unei ierarhii de stări cu conexiuni şi probabilităţi (în esenţă, un neocortex simulat pentru producerea vorbirii) şi apoi să folosim această reţea ierarhică determinată de stări pentru recunoaşterea noilor inputuri verbale. Pentru crearea unui sistem independent de vorbitor, puteam folosi mostre de la mai multe persoane diferite, pentru antrenarea Modelelor

Markov. Adăugând şi elementul ierarhizării pentru reprezentarea naturii ierarhice a informaţiei în limbă, le-am numit în consecinţă Modele Markov Ascunse Ierarhice, MMAI (*Hierarchical Hidden Markov Models* – HHMM).

Colegii mei din compania Kurzweil Applied Intelligence au fost sceptici în legătură cu funcţionarea acestei metode, dat fiind faptul că era o metodă autoorganizatoare, reminiscenţă a reţelelor neuronale, care nu mai erau la modă şi cu care avusesem prea puţin succes. Eu am scos în evidenţă faptul că reţeaua unui sistem neuronal în reţea este fixă şi nu se adaptează inputului: greutăţile se adaptează, dar nu şi conexiunile. În cazul sistemului bazat pe Modelul Markov, dacă acesta era conceput corect, sistemul urma să taie conexiunile nefolosite astfel încât, în esenţă, să adapteze topologia.

Am creat ceea ce se considera a fi un proiect „skunk works" (termen organizaţional aplicat unui proiect care se abate de la calea clasică şi dispune de puţine resurse formale), compus din mine, un programator part-time şi un inginer din domeniul electric (pentru crearea băncii de filtre de frecvenţe). Spre surpriza colegilor mei, efortul nostru a fost încununat de succes, sistemul reuşind să recunoască vorbirea pentru un vocabular larg, cu o mare acurateţe.

După acel experiment, toate eforturile noastre în domeniul recunoaşterii vorbirii s-au bazat pe Modelele Markov Ascunse Ierarhice. Se pare că şi alte



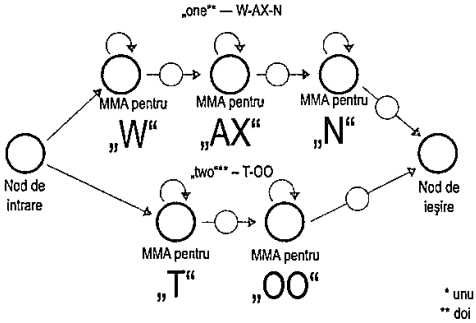
Stări Markov Ascunse şi posibile tranziţii pentru producerea unei succesiuni de cuvinte în text în limbaj natural companii preocupate de recunoaşterea vorbirii au reuşit să descopere în mod independent valoarea acestei metode şi, de la mijlocul anilor 1980, cea mai mare parte a

cercetărilor din domeniul recunoaşterii vorbirii automate s-au bazat pe această abordare. Modelele Markov Ascunse sunt folosite, de asemenea, în sinteza vorbirii – reţine, cititorule, că ierarhia neocorticală biologică este utilizată nu numai pentru recunoaşterea inputului, ci şi pentru producerea output-ului, de exemplu vorbirea şi mişcarea fizică.

Modelele Markov Ascunse Ierarhice, MMAI, sunt folosite şi în sisteme care înţeleg semnificaţia propoziţiilor în limbaj natural, ceea ce înseamnă urcarea în ierarhia conceptuală.

Pentru a înţelege cum funcţionează Modelele Markov Ascunse Ierarhice, MMAI, pornim de la o reţea care constă în toate tranziţiile de stare posibile. Metoda cuantificării vectoriale descrisă mai înainte este fundamentală aici, deoarece altminteri nu ar exista prea multe posibilităţi de luat în considerare.

Iată o topologie iniţială simplificată:



O topologie simplă pentru Modelul Markov Ascuns, pentru recunoaşterea a doua cuvinte rostite

Mostrele de rostire sunt procesate pe rând, câte una. Pentru fiecare, modificăm în mod recurent probabilităţile tranziţiilor, pentru a reflecta mai bine mostra de imput pe care tocmai am procesat-o. Modelele Markov folosite în recunoaşterea vorbirii codifică posibilitatea ca în fiecare fonem să se găsească anumite modele de sunete, modul în care se influenţează reciproc fonemele şi ordinea posibilă a fonemelor. Sistemul mai poate să cuprindă şi reţele de

probabilitate la niveluri superioare de structură a limbajului, cum ar fi ordinea cuvintelor, includerea de expresii şi aşa mai departe, urcând în ierarhia limbajului.

În vreme ce sistemele de recunoaştere anterioare încorporau reguli specifice referitoare la structurile fonemelor şi succesiuni explicit codificate de lingviştii umani, noului sistem bazat pe Modelele Markov Ascunse Ierarhice nu i s-a spus că în limba engleză există patruzeci şi patru de foneme, nu i s-au dezvăluit succesiunile de vectori probabile pentru fiecare fonem şi nici ce succesiuni de foneme sunt mai probabile decât altele. Am lăsat sistemul să descopere aceste „reguli" singur, din miile de ore de date de limbaj uman care au fost transcrise. Avantajul acestei abordări în raport cu regulile codificate manual este că modelele pun la punct reguli probabilistice de care experjii umani sunt rareori conştienţi. Am observat că multe dintre regulile pe care sistemul le-a învăţat în mod automat din date se deosebeau în moduri subtile, dar importante, de regulile stabilite de experţii umani.

Din momentul în care reţeaua a fost pregătită, am încercat să recunoaştem limbajul luând în considerare căi alternative în reţea şi alegând calea cea mai probabilă, dată fiind succesiunea respectivă de vectori de imput pe care o văzuserăm. Cu alte cuvinte, dacă vedeam o succesiune de stări care ar fi fost posibil să producă rostirea unui cuvânt, concluzionam că rostirea s-a produs în urma acelei succesiuni corticale. Această simulare de neocortex bazată pe Modelele Markov Ascunse Ierarhice conţinea cuvinte-etichetă, astfel încât era capabilă să propună o transcriere a ceea ce auzea.

Am putut pe urmă să ne perfecţionăm rezultatele şi mai mult, continuând să antrenăm reţeaua în timp ce o foloseam pentru recunoaştere. După cum am discutat, recunoaşterea şi învăţarea concomitente se petrec şi la toate nivelurile ierarhiei noastre neocorticale biologice.

Algoritme evoluţioniste (genetice)

Mai există un considerent foarte important: Cum

setăm numeroşii parametri care controlează funcţionarea unui sistem de recunoaştere a modelelor? Acestea pot să cuprindă numărul de vectori pe care îl permitem în pasul cuantificării vectoriale, topologia iniţială a stărilor ierarhice (înainte ca faza de antrenare a procesului Modelului Markov Ascuns să le reducă), pragul de recunoaştere de la fiecare nivel al ierarhiei, parametrii care controlează manevrarea parametrilor de mărime şi multe altele. Putem stabili toate acestea pe baza intuiţiei, dar rezultatele vor fi departe de a fi optime.

Numim aceşti parametri „Parametri-Dumnezeu" deoarece sunt stabiliţi înaintea metodei de determinare a topologiei Modelelor Markov Ascunse (sau, în cazul biologic, înainte ca persoana să-şi înveţe lecţiile prin crearea similară a conexiunilor în ierarhia sa corticală). Probabil că termenul este impropriu, dat fiind faptul că aceste detalii iniţiale de concepţie bazate pe ADN sunt determinate de evoluţia biologică, deşi unii pot să vadă în acest proces intervenţia mâinii lui Dumnezeu (şi, cu toate că personal consider că evoluţia este într-adevăr un proces spiritual, această discuţie îşi are locul în Capitolul 9).

Când am ajuns la setarea acestor „Parametri ai lui Dumnezeu" în sistemul nostru simulat ierarhic de învăţare şi recunoaştere, ne-am inspirat din nou de la natură şi am decis să-i perfecţionăm – în cazul nostru, folosind o simulare a evoluţiei. Am folosit algoritmii genetici sau evolutivi (AG) care includ reproducerea sexuală şi mutaţiile.

Iată o descriere simplificată a funcţionării acestei metode. Mai întâi, determinăm un mod de codificare a posibilelor soluţii pentru o problemă dată. Dacă problema este optimizarea parametrilor de concepţie pentru un circuit, atunci definim o listă cu toţi parametrii (cu un număr specific de biţi atribuit fiecărui parametru) care caracterizează circuitul. Această listă este considerată un cod genetic în algoritmul genetic. Pe urmă, generăm în mod aleatoriu mii de coduri genetice sau mai multe.

Fiecare asemenea cod genetic (care reprezintă un set de parametri de concepţie) este considerat un organism „soluţie” simulat.

Acum evaluăm fiecare organism simulat într-un mediu simulat, folosind o metodă bine definită de evaluare a fiecărui set de parametri. Evaluarea este o cheie a succesului pentru algoritmul genetic. În exemplul nostru, am rula fiecare program generat de parametri şi l-am judeca după criterii adecvate (şi-a îndeplinit sarcina, timpul de care a avut nevoie şi aşa mai departe). Celor mai bune organisme-soluţie (celor mai bune concepţii) li se permite să supravieţuiască, iar restul sunt eliminate.

Acum îi facem pe toţi supravieţuitorii să se înmulţească până când ajung la acelaşi număr de creaturi-soluţie. Această înmulţire se realizează prin simularea reproducerii sexuale: cu alte cuvinte, creăm noi progenituri care îşi extrag, fiecare, o parte din codul genetic de la un părinte, iar cealaltă parte de la al doilea părinte. De obicei nu se face nicio distincţie între organismele-mascul şi cele femelă; este suficient să se genereze o progenitură din oricare doi părinţi arbitrari, aşadar este vorba aici despre căsătorii între părţi de acelaşi sex. Probabil că mixajul nu este la fel de interesant ca reproducerea sexuată din lumea naturală, dar aspectul relevant stă în faptul că există doi părinţi. Pe măsură ce aceste organisme simulate se înmulţesc, permitem să apară anumite mutaţii (modificări aleatorii) la nivelul cromozomilor.

Am definit, aşadar, o generaţie de evoluţie simulată; acum repetăm aceşti paşi pentru fiecare generaţie ulterioară. La sfârşitul oricărei generaţii, determinăm măsura în care s-a perfecţionat concepţia (respectiv, computăm progresul mediu în funcţia de evaluare pentru toate organismele care au supravieţuit). Când gradul de progres în evaluarea creaturilor concepute de la o generaţie la alta devine foarte mic, oprim ciclul recurent şi folosim cea (cele) mai bună (bune) concepţie (concepţii)

din ultima generaţie. (Pentru o descriere a algoritmilor genetici, vezi această notă de final.) 11

Cheia unui algoritm genetic stă în faptul că creatorii lui umani nu programează o soluţie în mod direct; dimpotrivă, lăsăm să apară una printr-un proces recurent de competiţie şi progres simulate. Evoluţia biologică este inteligentă dar lentă, astfel încât, pentru a-i spori inteligenţa, îi accelerăm foarte mult ritmul greoi. Computerul este suficient de rapid pentru a simula mai multe generaţii într-un răstimp de câteva ore sau câteva zile; ocazional, le-am lăsat să ruleze săptămâni întregi, pentru a simula sute de mii de generaţii. Însă trebuie să trecem prin acest proces recurent o singură dată; de îndată ce am lăsat evoluţia simulată să-şi urmeze cursul, putem aplica regulile evoluate şi deosebit de rafinate la problemele reale într-o manieră foarte rapidă. Am folosit sistemele de recunoaştere a vorbirii pentru evoluarea topologiei iniţiale a reţelei şi a altor parametri fundamentali. Astfel, am folosit două metode de autoorganizare: un algoritm genetic pentru simularea evoluţiei biologice care a dat naştere unei anumite concepţii corticale şi Modele Markov Ascunse Ierarhice pentru simularea organizării corticale care însoţeşte învăţarea umană.

O altă cerinţă majoră pentru succesul unui algoritm genetic rezidă într-o metodă validă de evaluare a fiecărei soluţii posibile. Această evaluare trebuie realizată rapid, deoarece trebuie să ţină seamă de multe mii de posibile soluţii pentru fiecare generaţie a evoluţiei simulate. Algoritmii genetici sunt foarte abili în rezolvarea problemelor cu prea multe variabile, pentru care se cere să se computeze soluţii analitice precise. Conceperea unui motor, de exemplu, poate să implice mai bine de o sută de variabile şi impune satisfacerea câtorva zeci de constrângeri; algoritmii genetici folosiţi de cercetătorii de la General Electric au putut produce concepte de motoare cu reacţie care întrunesc constrângerile mult mai precis

decât metodele convenţionale.

În cazul folosirii algoritmilor genetici, trebuie, însă, să fii foarte atent la ceea ce anume ceri. Un algoritm genetic a fost folosit pentru rezolvarea problemei depozitării unor blocuri şi a dat o soluţie perfectă... cu excepţia faptului că presupunea mii de paşi. Programatorii umani au uitat să includă minimalizarea numărului de paşi în funcţia lor de evaluare.

Proiectul lui Scott Dave, *Electric Sheep*, este un algoritm genetic care produce lucrări de artă. Funcţia de evaluare foloseşte evaluatori umani în cadrul unei colaborări de tip *open-source* în care sunt implicaţi mii de oameni. Lucrările de artă se modifică în timp şi poţi să le vezi la adresa electricsheep.org.

Pentru recunoaşterea vorbirii, combinaţia dintre algoritmii genetici şi Modelele Markov Ascunse a funcţionat. extrem de bine. Simularea evoluţiei cu ajutorul unui algoritm genetic a putut să îmbunătăţească substanţial performanţa reţelelor pe bază de Modele Markov Ascunse Ierarhice. Ceea ce a adus evoluţia a fost net superior concepţiei noastre iniţiale, care se baza pe intuiţia noastră.

Pe urmă am experimentat introducerea unei serii de mici variaţii în sistemul de ansamblu. De exemplu, am produs perturbări (mici modificări aleatorii) ale inputului. O altă asemenea modificare a constat în a face Modelele Markov alăturate să se „scurgă" unele în altele, determinând rezultatele unui Model Markov să influenţeze modelele aflate în „apropiere".

Deşi nu ne-am dat seama la vremea respectivă, genul de ajustări pe care le-am experimentat seamănă foarte mult cu tipurile de modificări care se produc în structurile corticale biologice.

La început, asemenea schimbări afectează performanţa (măsurată în acurateţea recunoaşterii). Dar rerulând evoluţia (respectiv, reruiind algoritmul genetic) după operarea acestor alterări, sistemul s-a adaptat în

consecinţă, optimizându-se pentru modificările introduse. În general, procedura aceasta restaurează performanţa. Când, ulterior, am înlăturat schimbările introduse, performanţa s-a degradat, deoarece sistemul evoluase în direcţia compensării schimbărilor. Sistemul adaptat a devenit dependent de schimbări.

Un tip de alterare care a ajutat practic performanţa (după rerularea algoritmului genetic) a constat în introducerea unor mici modificări aleatorii în imput. Motivul acestei atitudini este binecunoscuta problemă a „overfitting”-ului care apare la sistemele autoorganizate. Există pericolul ca un asemenea sistem să suprageneralizeze în funcţie de exemplele specifice conţinute în mostra de pregătire. Operând ajustări aleatorii ale inputului, supravieţuiesc modelele invariabile din date şi prin asta sistemul învaţă aceste modele mai profunde. Metoda a fost de ajutor numai dacă am rerulat algoritmul genetic cu caracteristica de randomizare activată.

Aceasta introduce o dilemă în înţelegerea circuitelor noastre corticale biologice. S-a observat, de exemplu, că într-adevăr poate exista o slabă scurgere de la o conexiune corticală la alta, rezultată din modul în care se formează conexiunile biologice: se pare că proprietăţile electrochimice ale axonilor şi dendritelor sunt supuse efectelor electromagnetice ale conexiunilor din apropiere. Să presupunem că am putea face un experiment prin care să înlăturăm acest efect dintr-un creier real. Acest lucru ar fi dificil de realizat practic, dar nu neapărat imposibil. Să presupunem că am face un asemenea experiment şi am descoperi că circuitele corticale funcţionează mai puţin eficient în absenţa acestei scurgeri neuronale. Am putea concluziona, aşadar, că acest fenomen a reprezentat o intenţie foarte inteligentă a evoluţiei şi a avut o importanţă decisivă pentru ca neocortexul să poată ajunge la nivelul său specific de performanţă. Am putea scoate în evidenţă, mai mult, că un asemenea rezultat indică faptul că modelul

ordonat al fluxului ascendent de modele în ierarhia conceptuală şi fluxul descendent de predicţii în ierarhia conceptuală au fost, de fapt, complicate din cauza influenţei complexe dintre conexiuni.

Dar aceasta nu ar fi în mod necesar o concluzie exactă. Să ne gândim la experienţa noastră cu un cortex simulat bazat pe Modele Markov Ascunse Ierarhice, în care am implementat o modificare foarte asemănătoare cu încrucişarea interneuronală. Dacă am rerula evoluţia cu acest fenomen instituit, performanţa ar fi reinstaurată (deoarece procesul evoluţiei s-a adaptat la el). Dacă am înlocui apoi încrucişarea, performanţa ar fi din nou compromisă, în cazul biologic, evoluţia (respectiv evoluţia biologică) a „rulat" într-adevăr în condiţiile instaurării acestui fenomen. Parametri detaliaţi ai sistemului au fost setaţi, prin urmare, prin evoluţie biologică, pentru a deveni dependenţi de aceşti factori, astfel încât modificarea lor va afecta negativ performanţa dacă nu rulăm din nou evoluţia. Această abordare este fezabilă în lumea simulată, în care evoluţia durează numai câteva zile sau săptămâni, dar în lumea biologică ar necesita zeci de mii de ani.

Prin urmare, cum putem să ne dăm seama dacă o anumită caracteristică de proiectare a neocortexului biologic este sau nu o inovaţie vitală produsă de evoluţia biologică – respectiv una instrumentală pentru nivelul nostru de inteligenţă – sau pur şi simplu un artefact de care proiectarea sistemului este dependentă de-acum, dar fără de care nu ar fi putut să evolueze? Putem răspunde simplu la această întrebare rulând evoluţia simulată cu şi fără aceste variaţiuni particulare ale detaliilor de proiectare (de exemplu, cu şi fără conexiuni încrucişate). Putem face acest lucru chiar şi în cazul evoluţiei biologice, dacă examinăm evoluţia unei colonii de microorganisme la care generaţiile se măsoară în ore, însă nu este practică pentru organisme complexe ca omul. Acesta este un alt mare dezavantaj al biologiei.

Revenind la cercetările noastre din domeniul

recunoaşterii vorbirii, am descoperit că, dacă rulam evoluţia (respectiv un algoritm genetic) separat pe proiectul iniţial (1) al Modelelor Markov Ascunse Ierarhice care făceau modelele structurii interne a fonemelor şi (2) pe Modelele Markov Ascunse Ierarhice ale structurilor cuvintelor şi expresiilor, obţineam rezultate şi mai bune. Ambele niveluri ale sistemului foloseau Modele Markov Ascunse Ierarhice, dar algoritmul genetic asigura evoluţia unor variaţiuni de proiectare între aceste niveluri diferite. Această abordare permitea în continuare realizarea unor modele ale fenomenelor care se petrec între cele două niveluri, cum ar fi întinderea fonemelor care apare adeseori atunci când înşiruim anumite cuvinte unele după altele (de exemplu, *„How are you aii doing*?*”* poate să devină *„How re y all doing?”*).

Este posibil ca un fenomen similar să se fi manifestat în diferite regiuni ale. cortexului biologic, prin care au evoluat mici deosebiri bazate pe tipurile de modele de care se ocupau. În vreme ce toate aceste regiuni folosesc acelaşi algoritm neocortical fundamental, evoluţia biologică a avut suficient timp la dispoziţie pentru a face acordul fin al proiectului fiecăreia dintre ele, ca să devină optime pentru modelele lor. Pe de altă parte însă, aşa cum am arătat mai înainte, specialiştii în neuroştiinţă şi neurologii au observat în aceste zone o plasticitate substanţială, fapt ce sprijină ideea unui algoritm general neocortical. Dacă metodele fundamentale pentru fiecare regiune ar fi radical diferite, o asemenea interşanjabilitate între regiunile corticale nu ar fi posibilă.

Sistemele pe care le-am creat în cursul cercetărilor noastre folosind această combinaţie de metode de autoorganizare au fost încununate de un mare succes. În cazul recunoaşterii vorbirii, ele au reuşit pentru prima oară să facă faţă vorbirii continue normale şi unui vocabular relativ nerestricţionat. Am izbutit să atingem o acurateţe foarte înaltă pentru o mare varietate de vorbitori, accente şi dialecte. Actualul stadiu de vârf, în

timp ce scriu această carte, este reprezentat de un produs numit Dragon Naturally Speaking (Versión 11.5) pentru PC de la Nuance (anterior, Kurzweil Computer Products). Le sugerez tuturor să îl încerce dacă sunt sceptici în privinţa performanţelor sistemelor contemporane de recunoaştere a vorbirii – acurateţea este adeseori de 99 la sută sau chiar mai mare după câteva minute de antrenare pentru vocea respectivului vorbitor, la vocabulare relativ nerestricţionate. Dragon Dictation este o aplicaţie mai simplă, dar totuşi impresionantă pentru iPhone, şi nu a avut nevoie de antrenare vocală. Siri, asistentul personal al iPhone-urilor Apple contemporane, foloseşte aceeaşi tehnologie de recunoaştere a vorbirii, cu extensii menite a-l ajuta să facă faţă înţelegerii limbii naturale.

Performanţa acestor sisteme este o mărturie a forţei matematicii. Cu ajutorul lor, computăm practic ceea ce se petrece în neocortexul vorbitorului – chiar dacă nu avem acces direct la creierul persoanei respective – ca pas înainte vital în recunoaşterea spuselor unei persoane şi, în cazul unor sisteme ca Siri, a înţelesului mesajelor verbale respective. Am putea să ne întrebăm: *Dacă ne-am uita de-adevăratelea în neocortexul vorbitorului, am vedea conexiunile şi greutăţile corespunzătoare Modelelor Markov Ascunse Ierarhice compútate de software*? Aproape sigur, nu am găsi o potrivire precisă; structurile neuronale s-ar deosebi în mod invariabil de modelele din computer sub aspectul mai multor detalii. Cu toate acestea, aş susţine în continuare că trebuie să existe o echivalenţă matematică esenţială, la un grad înalt de precizie, între biologie şi încercarea noastră de a o imita; altminteri, aceste sisteme n-ar funcţiona atât de bine cum o fac.

LISP

LISP (LIST Processor) este un limbaj de computer propus iniţial de pionierul inteligenţei artificiale John McCarthy (1927 – 2011), în anul 1958. După cum sugerează şi denumirea, LISP se ocupă de liste. Fiecare

instrucţiune LISP este o listă de elemente; fiecare element este o altă listă sau un „atom”, care reprezintă un articol ireductibil, constituind fie un număr, fie un simbol. O listă inclusă într-o listă poate fi lista supremă, de unde rezultă că LISP este capabil de recurenţă. Instrucţiunile LISP pot fi recurente şi dacă lista include o listă şi aşa mai departe, până când este specificată lista originală. Pentru că listele pot să includă liste, LISP este capabil şi de procesare ierarhică. Lista poate fi condiţională, astfel încât să „emită” numai dacă elementele ei sunt satisfăcute. În acest fel, ierarhiile de asemenea condiţionalităţi pot fi folosite pentru identificarea calităţilor din ce în ce mai abstracte ale unui model.

LISP a ajuns un răcnet în comunitatea inteligenţei artificiale în anii 1970 şi la începutul anilor ’80. Fantezia pasionaţilor de LISP ai deceniului iniţial susţinea că limbajul acesta oglindeşte modul de funcţionare a creierului uman şi că orice proces inteligent poate fi codificat cel mai uşor şi mai eficient în LISP. A urmat un miniboom în companiile de „inteligenţă artificială” care ofereau interpreţi LISP şi produse LISP asociate, dar, la mijlocul anilor 1980, când a devenit evident că LISP, ca atare, nu era o scurtătură pentru crearea proceselor inteligente, bula investiţiilor a explodat.

S-a dovedit însă că pasionaţii LISP nu greşiseră întru totul. În esenţă, toţi recognizatorii de modele din neocortex pot fi consideraţi instrucţiuni LISP – fiecare dintre ei constituind o listă de elemente şi fiecare element putând fi o altă listă. Neocortexul este, prin urmare, angajat în procesarea listelor de o natură simbolică, similar cu ceea ce se petrece într-un program LISP. Mai mult de-atât, procesează toate cele 300 de milioane de „instrucţiuni” gen LISP simultan.

Pe de altă parte, din lumea LISP lipseau două caracteristici importante, dintre care una era învăţarea. Trebuia ca programele LISP să fie codificate linie cu linie de programatori umani. Au existat tentative de codificare

automată a acestor programe, folosindu-se o varietate de metode, dar acestea nu au făcut parte integrantă din conceptul limbajului. Neocortexul, în schimb, se programează pe el însuşi, umpăindu-şi „declaraţiile” (respectiv listele) cu informaţii semnificative şi executabile din propria experienţă şi din propriile bucle de feedback. Acesta este un principiu-cheie al modului de funcţionare a neocortexului: fiecare dintre recognizatorii săi de modele (respectiv fiecare instrucţiune LISP) este capabil (a) să îşi realizeze propria listă şi să se conecteze atât în sus, cât şi în jos la alte liste. A doua diferenţă rezidă în mărimea parametrilor. Se poate crea o variantă de LISP (codificată în LISP) care să permită. manevrarea unor asemenea parametri, dar această variantă nu face parte din limbajul de bază.

LISP este consecvent filosofiei originale a domeniului inteligenţei artificiale, care presupune găsirea de soluţii inteligente la probleme şi codificarea acestor soluţii direct în limbajul-maşină. Prima tentativă de creare a unei metode de autoorganizare care să înveţe din propria experienţă – reţelele neuronale – n-a avut succes, deoarece n-a asigurat un mijloc pentru modificarea topologiei sistemului ca răspuns la învăţare. Modelele Markov Ascunse Ierarhice au asigurat asta în mod eficient prin mecanismul lor de tăiere. Astăzi, Modelele Markov Ascunse Ierarhice împreună cu verii lor matematici constituie elemente majore în lumea inteligenţei artificiale.

Un corolar al observării similitudinilor dintre LISP şi structura de listă a neocortexului reprezintă un argument adus de cei ce insistă că, totuşi, creierul este prea complicat pentru ca noi să-l putem înţelege. Criticii scot în evidenţă faptul că un creier are miliarde de conexiuni şi, de vreme ce fiecare dintre ele trebuie să se afle acolo potrivit proiectului, ele constituie echivalentul a mii de miliarde de linii de cod. Aşa cum am văzut, am estimat că la nivelul neocortexului există ceva de ordinul a 300 de milioane de procesori de modele – sau 300 de milioane de

liste, cu proprietatea că fiecare element dintr-o listă indică o altă listă (sau, la nivelul conceptual cel mai redus, un model de bază ireductibil din afara neocortexului). Dar cifra de 300 de milioane este una rezonabil de mare de instrucţiuni LISP şi e într-adevăr mai mare decât orice program scris de om şi aflat în funcţiune.

În orice caz, trebuie să reţinem că aceste liste nu sunt practic specificate în concepţia iniţială a sistemului nervos. Creierul le creează din propriile sale experienţe. Acesta este secretul-cheie ai neocortexului. Procesele care realizează această autoorganizare sunt mult mai simple decât cele 300 de milioane de instrucţiuni care constituie capacitatea neocortexului. Aceste procese sunt specificate în genom. Aşa cum voi demonstra în Capitolul 11, volumul de informaţie unică din genom (după comprimarea fără pierdere) aplicată creierului este de aproximativ 25 de milioane de biţi – echivalentul a mai puţin de un milion de rânduri de cod. Complexitatea algoritmică propriu-zisă este chiar mai mică de-atât, pentru că, din cele 25 de milioane de biţi de informaţie genetică, partea cea mai însemnată se referă la nevoile biologice ale neuronilor şi nu neapărat la capacitatea lor de procesare a informaţiei. Însă chiar şi 25 de milioane de biţi de informaţie de concepţie reprezintă un nivel de complexitate căruia putem să-i facem faţă.

Sisteme ierarhice de memorie

După cum am spus în Capitolul 3, Jeff Hawkins şi Dileep George au pus la punct în anii ’2003 şi 2004 un model de neocortex care încorporează listele ierarhice descrise în cartea din 2004 a lui Hawkins şi Blakeslee *On Intelligence – Despre inteligenţă*. *O* prezentare mai actualizată şi foarte elegantă a memoriei temporale ierarhice poate fi găsită în dizertaţia de doctorat din anul 2008 a lui Dileep George.12 Numenta a implementat-o într-un sistem numit NuPIC (Numenta Platform for Intelligent Computing – Platforma Numenta pentru computaţie inteligentă) şi a pus la punct sisteme de

recunoaştere a modelelor şi de căutare de date pentru clienţi cum ar fi Forbes şi Power Analytics Corporation. După ce a lucrat la Numenta, George a fondat o nouă companie, intitulată Vicarious Sistems, cu finanţare de la Founder Fund (condusă de Peter Thiel, investitorul care a stat în spatele Facebook, şi de Sean Parker, primul preşedinte Facebook) şi de la Good Ventures, condusă de Dustin Maskovitz, cofondator Facebook. George a comunicat progrese semnificative în modelarea, învăţarea şi recunoaşterea automată a informaţiilor cu un număr substanţial de ierarhii. Şi-a numit sistemul „reţea corticală recursivă" şi intenţionează să realizeze aplicaţii pentru imagistica şi robotica medicală, dar nu numai. Tehnica Modelelor Markov Ascunse Ierarhice este foarte asemănătoare din punct de vedere matematic cu aceste sisteme de memorie ierarhică, mai ales dacă permitem sistemului bazat pe Modele Markov Ascunse Ierarhice să îşi organizeze propriile conexiuni între modulele de recunoaştere a modelelor. După cum am menţionat anterior, Modelele Markov Ascunse Ierarhice mai asigură un element adiţional important, care este modelarea distribuţiei aşteptate a magnitudinii (într-un continuum oarecare) pentru fiecare imput la computarea probabilităţii existenţei modelului avut în vedere. Recent, am lansat o nouă companie, denumită Patterns, Inc., care intenţionează să dezvolte modele neocorticale autoorganizate ierarhic pe baza Modelelor Markov Ascunse Ierarhice şi tehnici asociate în scopul înţelegerii limbajului natural. Un accent important va fi pus pe abilitatea sistemului de a-şi concepe propriile ierarhii într-o manieră similară neocortexului biologic. Sistemul avut în vedere de noi va citi în mod continuu o vastă gamă de materiale, cum ar fi Wikipedia şi alte resurse de cunoştinţe, ascultând de asemenea tot ce i se spune şi urmărind tot ce se scrie (dacă i se permite). Scopul este să devină un prieten util, care să-ţi răspundă la întrebări – chiar dinainte să le formulezi – şi să-ţi dea informaţii şi

ponturi utile în viaţa de zi cu zi.

Frontiera în mişcare a inteligenţei artificiale: *ascensiunea în ierarhia competenţei*

1. Un discurs lung şi plictisitor ţinut de un *topping* spumos de plăcintă.

2. Articol de îmbrăcăminte purtat de un copil, eventual la bordul unui vas de operă.

3. Căutat de doisprezece ani pentru crima de a-i fi mâncat pe războinicii regelui Hrothgar; cazul îi este dat ofiţerului Beowulf.

4. Poate să însemne să dezvolţi treptat în minte sau să porţi în timpul sarcinii.

5. Ziua Naţională a învăţătorului şi Ziua Derby-ului din Kentucky.

6. Wordsworth a spus că ei se avântă, dar niciodată nu hoinăresc.

7. Cuvânt din patru litere pentru fierul care se pune pe copita calului sau pe masa de joc dintr-un cazinou.

8. În actul trei al operei lui Verdi din anul 1846, acest Bici al lui Dumnezeu este înjunghiat şi omorât de iubita lui, Odabella.

— Exemple de întrebări *Jeopardy*! la toate Watson dând răspunsuri corecte.

Răspunsurile sunt: meringue harangue Doc de cuvinte intraductibil]; şorţ; Grendel; gestaţie; mai; ciocârlie; potcoavă.

La a treia întrebare, Watson a răspuns:

„Ce este Attila?” Gazda a replicat: „Fii mai exact”. Watson a clarificat cu „Ce este Attila Hunul?”, ceea ce este corect.

Tehnicile computerului pentru dezlegarea indiciilor la *Jeopardy*! seamănă cu ale mele. Maşina se concentrează pe cuvintele-cheie din indiciu, apoi îşi consultă memoria (în cazul lui Watson, o bancă de date de 15 terabiţi de cunoştinţe umane) în căutarea unor aglomerări de asociaţii cu aceste cuvinte. Confruntă cu rigurozitate principalele răspunsuri cu informaţiile contextuale pe care le poate

găsi: denumirea categoriei; genul de răspuns care e aşteptat; momentul, locul şi genul masculin sau feminin sugerate de indiciu şi aşa mai departe. Iar când se simte suficient de „sigur”, decide să apese soneria. Tot acest proces este instantaneu, intuitiv pentru un jucător uman la *Jeopardy*! dar eu am fost convins că, sub glugă, şi creierul meu făcea mai mult sau mai puţin acelaşi lucru.

— Ken Jennings, campion uman la *Jeopardy*!

care a pierdut în faţa lui Watson

Eu, unul, îi întâmpin cu braţele deschise pe noii noştri stăpâni roboţi.

— Ken Jennings, parafrazându-i pe Simpsoni, după ce a pierdut în faţa lui Watson

Dumnezeule mare! [Watson] este mai inteligent decât jucătorul mediu de *Jeopardy*! *în* ceea ce priveşte răspunsurile la întrebările de la *Jeopardy*!

Asta înseamnă impresionant de inteligent.

— Sebastián Thrun, fost director al Laboratorului de Inteligenţă

Artificială de la Stanford

Watson nu înţelege nimic. E doar un buldozer mai mare.

— Noam Chomsky

Inteligenţa artificială există pretutindeni în jurul nostru – nu mai avem niciun cuvânt de spus în privinţa asta. Simplul gest al conectării cu cineva prin intermediul unui mesaj de tip text, e-mail sau apel pe un telefon mobil foloseşte algoritmi inteligenţi pentru rutarea informaţiei. Aproape toate produsele pe care le atingem sunt proiectate iniţial printr-o colaborare între o inteligenţă artificială şi una umană, iar apoi sunt realizate în fabrici automatizate. Dacă toate sistemele de inteligenţă artificială ar decide ca mâine să intre în grevă, civilizaţia noastră ar fi paralizată: nu am mai putea scoate bani de la bancă şi până şi banii ar dispărea; comunicaţiile, transportul şi producţia s-ar opri brusc. Din fericire, maşinile noastre inteligente nu sunt deocamdată atât de

inteligente încât să organizeze o asemenea conspiraţie.

Nouă astăzi în inteligenţa artificială este natura visceral impresionantă a exemplelor disponibile în spaţiul public. Bunăoară, să ne gândim la automobilele Google care. Se conduc singure (în momentul în care scriu aceste rânduri, ele au trecut de 320.000 de kilometri parcurşi în oraşe mari şi mici), o tehnologie ce va conduce la un număr semnificativ redus de accidente rutiere, la creşterea capacităţii drumurilor, la eliminarea necesităţii ca omul să îşi asume povara şofatului şi la multe alte beneficii. Deja, automobilele fără şofer sunt, practic, legale pe drumurile publice din Nevada, deşi sunt supuse unor restricţii, dar utilizarea lor de către publicul larg din întreaga lume nu este aşteptată înaintea următorului deceniu. Tehnologia care urmăreşte drumul în mod inteligent şi îl avertizează pe şofer cu privire la pericolele iminente este deja instalată pe automobile. O asemenea tehnologie se bazează parţial pe modelul de succes al procesării vizuale la nivelul creierului, creată de Tomaso Poggio de la MIT. Denumită Mobileye, ea a fost pusă la punct de Amnon Shashuna, fost doctorand al lui Poggio. Tehnologia este capabilă să îl avertizeze pe şofer cu privire la pericole cum ar fi o coliziune iminentă sau un copil care aleargă în calea automobilului şi a fost utilizată recent de producători precum Volvo şi BMW.

În această secţiune a cărţii mă voi concentra asupra tehnologiilor limbajului – şi asta din mai multe motive. Deloc surprinzător, natura ierarhică a limbajului oglindeşte fidel natura ierarhică a gândirii noastre. Limbajul vorbit a fost prima noastră tehnologie, iar limbajul scris – cea de-a doua. Propriile mele cercetări în domeniul inteligenţei artificiale, după cum a demonstrat acest capitol, s-au concentrat puternic pe limbaj. În cele din urmă, stăpânirea limbajului este o capacitate cu o extraordinară forţă de influenţă. Watson a citit deja sute de milioane de pagini pe web şi stăpâneşte toate cunoştinţele conţinute în aceste documente. În cele din urmă, maşinile vor fi capabile să

acumuleze întreaga cunoaştere de pe web – care, în esenţă, reprezintă toată cunoaşterea civilizaţiei om-maşină.

Matematicianul englez Alan Turing (1912 – 1954) şi-a întemeiat testul eponim pe capacitatea computerului de a conversa în limbaj natural folosind mesaje text.13 Turing a considerat că întreaga inteligenţă umană a fost întrupată şi reprezentată de limbaj şi că nicio maşină nu poate trece Testul Turing prin simple trucuri de limbaj. Deşi Testul Turing este un joc care implică limbajul scris, Turing credea că singurul mod în care poate să-l treacă un computer este ca maşina să posede într-adevăr echivalentul inteligenţei de nivel uman. Criticii au avansat ipoteza că un adevărat test al inteligenţei de nivel uman ar trebui să includă şi stăpânirea informaţiei vizuale şi auditive.14 Deoarece multe dintre propriile mele proiecte ce ţin de inteligenţa artificială presupun învăţarea computerului să stăpânească informaţii senzoriale cum ar fi vorbirea umană, formele literelor şi sunetele muzicale, s-ar aştepta din partea mea să fiu partizanul includerii acestor tipuri de informaţie într-un test adevărat de inteligenţă. Totuşi, sunt de acord cu intuiţia iniţială a lui Turing că versiunea exclusiv text a Testului Turing este suficientă. Adăugarea inputului vizual sau auditiv nu ar face cu adevărat testul mai greu de trecut.

Nu trebuie să fii expert în inteligenţă artificială ca să fii impresionat de performanţa lui Watson la *Jeopardy*! Cu toate că înţeleg la un nivel rezonabil metodologia folosită la multe dintre subsistemele sale cheie, asta nu-mi diminuează reacţia emoţională atunci când îl văd în acţiune. Nici măcar o înţelegere perfectă a modului de funcţionare a tuturor sistemelor lui componente – la care nu a ajuns încă nimeni – nu te-ar putea ajuta să anticipezi felul cum va reacţiona de fapt Watson la o situaţie dată. El conţine sute de subsisteme care interacţionează şi fiecare dintre acestea analizează milioane de ipoteze concurente în acelaşi timp, deci anticiparea rezultatului este

imposibilă. O analiză temeinică – *post-factum* – a deliberărilor lui Watson pentru o singură căutare de trei secunde ar necesita secole de viaţă de om.

Ca să-mi continuu propria istorie, la sfârşitul anilor 1980 şi începutul anilor ’90, am început să lucrăm la înţelegerea limbajului natural în câteva domenii limitate. Poţi să-i vorbeşti unuia dintre produsele noastre, denumit Kurzweil Voice, despre orice, câtă vreme subiectul are legătură cu editarea documentelor. (De exemplu, „Mută al treilea paragraf de la pagina anterioară aici.”) Produsul a dat rezultate destul de bune în acest domeniu limitat dar util. Am creat de asemenea sisteme cu cunoştinţe din domeniul medical, astfel încât medicii să-şi poată dicta rapoartele despre pacienţi. Sistemul nostru avea cunoştinţe suficiente din domenii cum ar fi radiologia şi patologia ca să-i poată adresa întrebări medicului dacă ceva din raport i se părea neclar şi putea să-l călăuzească pe acesta în procesul de raportare. Sistemele de raportare medicală au evoluat, devenind o afacere de miliarde de dolari la Nuance.

Înţelegerea limbajului natural, în special ca extensie a recunoaşterii automate a vorbirii, a devenit un fapt cotidian. În momentul în care scriu această carte, Siri, asistentul personal automat de pe iPhone 4S, produce furori în lumea computaţiei mobile. Îi poţi cere lui Siri să facă orice e capabil să facă un smartphone care se respectă (de exemplu: „Unde pot să găsesc aici, primprejur, mâncare indiană?” sau „Transmite-i un mesaj text soţiei că sunt pe drum” sau „Ce părere are lumea de noul film al lui Brad Pitt?”) şi, de cele mai multe ori, Siri se va conforma. Siri poate susţine şi o conversaţie redusă nonproductivă. Dacă întrebi care este sensul vieţii, îţi va răspunde „42”, replică în care pasionaţii *The Hitchhiker’s Guide to the Galaxy-Ghidul autostopistului în galaxie* vor recunoaşte „răspunsul la întrebarea supremă despre viaţă, Univers, tot”. La întrebările care ţin de cunoaştere (inclusiv cea despre sensul vieţii) răspunde Wolfram Alphy,

descris ia pagina 164. Există o întreagă lume de „chatboţi” care nu fac nimic altceva în afară de a se angaja în conversaţii mărunte. Dacă vrei să vorbeşti cu chatbotul nostru pe nume Ramona, du-te ia website-ul KurzweilAI.net şi dă clic pe „Chat cu Ramona”.

Unii oameni mi s-au plâns în legătură cu incapacitatea lui Siri de a răspunde la anumite comenzi, dar îmi amintesc de multe ori că este vorba despre aceiaşi oameni care se plâng consecvent şi de oamenii care asigură servicii. Uneori sugerez să încercăm împreună şi adeseori funcţionează mai bine decât se aşteaptă aceşti oameni. Plângerile îmi amintesc de povestea cu câinele care joacă şah. Unui neîncrezător care îi întreabă, stăpânul câinelui îi răspunde: „Da, e adevărat, chiar joacă şah, dar are finalurile slabe”, încep să apară concurenţi eficienţi, cum ar fi Google Voice Search.

Faptul că publicul face acum conversaţie în limbaj natural vorbit cu computerele ţinute în mână marchează o nouă eră. Este ceva obişnuit ca oamenii să respingă semnificaţia unei tehnologii de primă generaţie din cauza limitărilor ei. După vreo câţiva ani, când tehnologia funcţionează bine, oamenii încă îi mai resping importanţa pentru că, asta este, nu mai e nouă. Acestea fiind spuse, Siri funcţionează impresionant pentru un produs de primă generaţie şi este limpede că această categorie de produse se va îmbunătăţi din ce în ce.

Siri foloseşte tehnologiile de recunoaştere a vorbirii bazate pe Modele Markov Ierarhice de la Nuance. Extensiile în limbaj natural au fost puse la punct prima oară de proiectul intitulat „CALO”, finanţat de DARPA.15 Siri a fost îmbunătăţit cu tehnologiile pentru limbajul natural ale Nuance, care oferă şi o tehnologie foarte asemănătoare intitulată Dragon Go! 16

Metodele folosite pentru înţelegerea limbii naturale sunt similare cu Modelele Markov Ascunse şi foarte des sunt folosite Modelele Markov Ascunse Ierarhice. În vreme ce linele dintre aceste sisteme nu sunt etichetate precis ca

folosind Modele Markov Ascunse sau Modele Markov Ascunse Ierarhice, suportul matematic este practic identic. Toate implică ierarhii de succesiuni liniare în care fiecare element are o greutate, conexiuni autoadaptative şi un sistem de ansamblu autoorganizat, bazat pe învăţarea datelor. De obicei, învăţarea continuă în cursul utilizării propriu-zise a sistemului. Această abordare corespunde structurii ierarhice a limbii naturale – este doar o extensie naturală a scării conceptuale de la părţi de vorbire, la cuvinte, expresii şi structuri semantice. Ar fi logic să se ruleze un algoritm genetic pe aceşti parametri care controlează exact algoritmul de învăţare specific acestei clase de sisteme de învăţare Ierarhice şi care determină detaliile optime ale algoritmului.

Pe parcursul ultimului deceniu, s-a înregistrat o modificare a felului în care sunt create structurile ierarhice. În anul 1984, Douglas Lenat (născut în 1950) a început lucrul la ambiţiosul Proiect Cyc (de la enl YCI – opedic), care urmărea să creeze reguli capabile să codifice cunoştinţele cotidiene „de bun-simţ. Aceste reguli erau organizate într-o colosală ierarhie şi fiecare implica – din, nou – o succesiune liniară de stări. De exemplu, o regulă Cyc poate spune că un câine are un chip. Pe urmă, Cyc se poate conecta la regulile generale privitoare la structura chipurilor: un chip are doi ochi, un nas, o gură şi aşa mai departe. Nu trebuie să avem un set de reguli pentru chipul unui câine şi altul pentru chipul unei pisici, cu toate că, desigur, ne-am putea dori să includem reguli suplimentare privitoare la felul în care se deosebeşte chipul câinelui de cel al pisicii. Sistemul include, de asemenea, un motor de interferenţă: dacă avem reguli care afirmă că un cocker Spaniei este câine, câinii sunt animale, iar animalele se hrănesc şi dacă am întreba motorul de interferenţă dacă cocker spanielii mănâncă, sistemul ar răspunde că da, cocker spanielii mănâncă. Pe parcursul următorilor douăzeci de ani şi după mii de ani-persoană de eforturi, au fost scrise şi testate peste un milion de asemenea reguli.

Fapt interesant, limbajul pentru scrierea regulilor Cyc – denumit Cyel – este aproape identic cu LISP.

Între timp, o şcoală de gândire opusă considera că abordarea cea mai bună pentru înţelegerea limbii naturale şi pentru crearea sistemelor inteligente în general se bazează pe învăţarea automată prin contactul zilnic cu un număr foarte mare de ipostaze ale fenomenului pe care încerca să-l stăpânească sistemul. Un exemplu elocvent al unui asemenea sistem este Google Translate, care poate să traducă în şi din cincizeci de limbi. Asta înseamnă 2.500 de direcţii de traducere diferite, deşi, pentru majoritatea perechilor de limbi, în loc să se facă traducerea directă a limbii 1 în limba 2, sistemul traduce limba 1 în engleză şi apoi engleza în limba 2. Acest fapt reduce numărul de translatoare Google care trebuie construite la nouăzeci şi opt (plus un număr de perechi nonengleză pentru care există traducere directă). Translatoarele Google nu folosesc reguli gramaticale; dimpotrivă, ele creează vaste baze de date pentru fiecare pereche de limbi cu traduceri frecvente pe baza unui corpus mare în genul „Piatra Rosetta” de documente traduse între aceste două limbi. Pentru cele şase limbi care constituie limbile oficiale ale Naţiunilor Unite, Google a utilizat documente ale Naţiunilor Unite, aşa cum sunt publicate acestea în fiecare dintre cele şase limbi. În cazul limbilor mai puţin obişnuite, au fost folosite alte idiomuri. Rezultatele sunt adeseori impresionante. DARPA organizează concursuri anuale pentru cele mai bune sisteme de traducere automată în şi din diferite perechi de limbi, iar Google Translate câştigă adeseori pentru anumite perechi, depăşind sisteme create direct de lingvişti umani.

Pe parcursul ultimului deceniu, două intuiţii majore au influenţat profund domeniul înţelegerii limbilor naturale. Prima este legată de ierarhii. Cu toate că abordarea Google a debutat cu asocierea succesiunilor plate de cuvinte dintr-o limbă în alta, natura inerent ierarhică a limbii s-a insinuat în mod inevitabil în această

operaţie. Sisteme care încorporează metodic învăţarea ierarhică (aşa cum sunt Modelele Markov Ascunse Ierarhice) au asigurat o performanţă semnificativ mai bună. Pe de altă parte însă, asemenea sisteme nu se construiesc chiar automat. Oamenii trebuie să înveţe aproximativ o ierarhie conceptuală o dată şi acelaşi lucru este valabil şi pentru sistemele computerizate, astfel încât procesul de învăţare trebuie gestionat cu grijă.

Cealaltă intuiţie este că regulile manuale funcţionează foarte bine în cazul unui miez de cunoştinţe comune de bază. Pentru traducerea pasajelor scurte, această abordare dă adeseori rezultate precise. De exemplu, DARPA a clasificat translatoarele pe bază de reguli din chineză în engleză mai bine decât Google Translate pentru pasaje scurte. Pentru ceea ce se numeşte trena limbii, care se referă la milioane de expresii ocazionale şi de concepte folosite în ele, acurateţea sistemelor bazate pe reguli se apropie de o asimptotă inacceptabil de coborâtă. Dacă plotăm acurateţea înţelegerii limbii naturale raportată la volumul de date de pregătire utilizate, constatăm că sistemele bazate pe date înregistrează iniţial performanţe superioare, dar scad la o acurateţe medie de circa 70 la sută. Într-un contrast acut, sistemele statistice pot ajunge până la 90 la sută în privinţa acurateţei, dar necesită un volum foarte mare de date pentru atingerea acestui nivel.

Adeseori avem nevoie pentru pregătire de o combinaţie formată dintr-o performanţă cel puţin moderată la un nivel redus de date şi de oportunitatea de a ajunge la o acurateţe ridicată cu un volum mai mare. Atingerea rapidă a unei performanţe moderate ne permite să scoatem pe piaţă un sistem, ulterior acesta adunând automat date de pregătire, pe măsură ce este folosit de oameni. În acest fel, se poate atinge un nivel foarte ridicat de învăţare în intervalul de timp în care este folosit sistemul, iar acurateţea acestuia se îmbunătăţeşte. Învăţarea statistică trebuie să fie pe deplin ierarhică

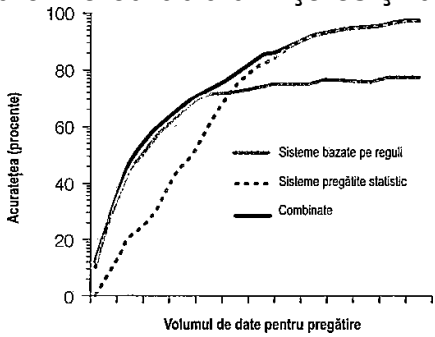
pentru a reflecta natura limbii, care oglindeşte, la rându-i, modul de funcţionare a creierului uman.

Aşa funcţionează Siri şi Dragon Gol – folosind reguli pentru fenomenele cele mai obişnuite şi sigure şi învăţând apoi „trena" limbii atunci când ajung în mâinile utilizatorilor reali. Când şi-a dat seama că a atins un plafon de performanţă bazat pe reguli codificate manual, echipa Cyc a adoptat şi ea această abordare. Regulile codificate manual asigură două funcţii esenţiale. Prima: oferă o acurateţe Iniţială adecvată, astfel încât un sistem în faza de încercări poate fi setat pentru utilizare largă, situaţie în care se va îmbunătăţi automat. A doua: asigură o bază solidă pentru nivelurile inferioare ale ierarhiei conceptuale, astfel încât învăţarea automată poate să înceapă de la niveluri conceptuale mai înalte.

După cum am menţionat anterior, Watson constituie un exemplu impresionant de combinare a regulilor codificate manual cu învăţarea statistică ierarhică. IBM a mixat mai multe programe de prim rang de limbă naturală pentru crearea unui sistem care să joace *Jeopardy*! în limbaj natural! în intervalul 14 – 16 februarie 2011, Watson a concurat cu doi dintre jucătorii umani de vârf: Brad Rutter, care a câştigat mai mulţi bani decât oricine altcineva la acest concurs de televiziune cu întrebări şi răspunsuri şi Ken Jennings, care anterior participase la campionatul de *Jeopardy*! pe durata record de şaptezeci şi cinci de zile.

În cadrul contextului, în prima mea carte, *The Age of Intelligent Machines* (*Epoca maşinilor inteligente*, Editura Paralela 45), scrisă la mijlocul anilor 1980, am anticipat că, înainte de 1998, un computer va câştiga campionatul mondial de şah. Am mai anticipat (1) că, atunci când se va întâmpla acest lucru, fie ne vom retrograda opinia cu privire la inteligenţa umană, fie ne-o vom îmbunătăţi pe cea privitoare la inteligenţa maşinii, fie vom minimaliza importanţa şahului şi (2) că, dacă ne va fi călăuză istoria, vom minimaliza şahul. Ambele lucruri s-au întâmplat în

anul 1997. Când supercomputerul jucător de şah produs de IBM şi denumit Deep Blue l-a înfrânt pe campionul mondial la şah Garry Kasparov, am fost întâmpinaţi imediat cu argumentul că era de aşteptat ca un computer să câştige, deoarece computerele sunt maşini logice, iar şahul, în ultimă instanţă, este un joc de logică. Astfel, victoria lui Deep Blue a primit sentinţa că nu este nici surprinzătoare şi nici semnificativă. Mulţi dintre criticii maşinii au mers mai departe şi au susţinut că un computer nu va putea stăpâni niciodată subtilităţile limbii umane, inclusiv metaforele, sinonimele, jocurile de cuvinte, formulările cu dublu înţeles şi umorul.



Acurateţea sistemelor de înţelegere a limbii naturale vorbite, exprimată în funcţie de volumul de date pentru pregătire. Cea mai bună abordare este să se combine regulile pentru „miezul” limbii cu o abordare bazată pe date pentru „trena* limbii

Acesta este cel puţin un motiv pentru care Watson reprezintă un reper atât de important: *Jeopardy*! este o sarcină sofisticată şi plină de provocări legate de limbă. În mod curent, întrebările *Jeopardy*! includ multe dintre fanteziile limbii umane. Ceea ce poate nu devine atât de evident pentru mulţi observatori este că, pe lângă obligaţia lui Watson de a stăpâni limba în contextul întrebărilor neaşteptate şi întortocheate, în ecuaţie a intervenit şi faptul că, în marea lor parte, cunoştinţele lui nu au fost codificate manual. Watson deţinea aceste cunoştinţe după ce citise practic 200 de milioane de pagini de documente în limbă naturală, inclusiv toată *Wikipedia* şi alte

enciclopedii, acumulând 4 mii de miliarde de biţi de cunoştinţe bazate pe limbă. După cum bine ştie cititorul acestei cărţi, *Wikipedia* nu e scrisă în LISP sau Cyel, ci în propoziţii naturale, care conţin toate ambiguităţile şi complexitatea inerente limbii. Watson a trebuit să consulte toate cele 4 mii de miliarde de caractere din materialul lui de referinţă pentru a răspunde la o întrebare. (îmi dau seama că „întrebările" de la *Jeopardy*! sunt de fapt răspunsuri care îşi caută întrebările, dar acesta este un aspect tehnic – şi, în ultimă instanţă, ele sunt totuşi întrebări.) Dacă Watson poate înţelege şi răspunde la întrebări în trei secunde pe baza informaţiilor cuprinse în 200 de milioane de pagini, nimic nu poate opri sisteme similare să citească celelalte miliarde de documente de pe web. Şi, într-adevăr, această operaţiune este în curs de desfăşurare chiar acum [în momentul scrierii cărţii – n.trad.].

În perioada în care puneam la punct sistemele de recunoaştere a caracterelor şi vorbirii şi sistemele-pionier de înţelegere a limbii naturale, din anii 1970 şi până în anii 1990, am folosit o metodologie de încorporare a unui „manager-expert". Puneam la punct mai multe sisteme care făceau acelaşi lucru, dar în fiecare dintre ele încorporam mici abordări diferite. Unele diferenţe erau subtile, ca de pildă variaţii ale parametrilor care controlează latura matematică a algoritmului învăţării. Alte diferenţe erau fundamentale, ca de exemplu includerea sistemelor bazate pe reguli în locul sistemelor de învăţare ierarhic-statistică. Managerul-expert era în sine un program software conceput pentru a învăţa punctele-forte şi punctele slabe ale diferitelor sisteme prin examinarea performanţelor acestora în situaţii ţinând de lumea reală. El se baza pe ideea că aceste puncte-forte erau ortogonale; respectiv, un sistem tindea să fie forte acolo unde altul tindea să fie slab. Şi, într-adevăr, performanţa însumată a unor asemenea sisteme combinate cu un manager-expert aflat la conducere a fost cu mult mai

mare decât performanţele oricăruia dintre sistemele individuale.

Watson funcţionează la fel. Folosind o arhitectură denumită URMA (Unstructured Information Managementărchitecture-Arhitectură nestructurată de gestionare a informaţiei), Watson conţine literalmente sute de sisteme diferite – multe dintre componentele sale de limbi individuale sunt identice cu cele folosite pentru sistemele de înţelegere a limbii naturale disponibile în mod public – toate încercând fie să dea un răspuns direct la o întrebare de *Jeopardy*! fie, cel puţin, să realizeze o anumită clarificare a întrebării. URMA funcţionează, în esenţă, ca un manager-expert pentru combinarea inteligentă a rezultatelor sistemelor independente şi depăşeşte cu mult sistemele anterioare – de felul celui pe care l-am pus noi la punct la compania predecesoare Nuance – prin faptul că sistemele lui individuale pot contribui la rezultat fără să dea în mod necesar răspunsul final. Este suficient ca un subsistem să ajute la îngustarea domeniului soluţiei. De asemenea, URMA e capabilă să computeze măsura în care are încredere în răspunsul final. Şi creierul uman face acest lucru – probabil că suntem foarte încrezători în răspunsul nostru atunci când ni se cere numele mic al mamei, dar nu suntem la fel de încrezători în privinţa numelui unei persoane întâlnite ocazional în urmă cu un an.

Prin urmare, în loc să ofere o singură abordare elegantă pentru înţelegerea problemei de limbă inerentă în *Jeopardy*! oamenii de ştiinţă de la IBM au combinat toate modulele de top în înţelegerea limbii pe care au putut să le găsească. Unele folosesc Modele Markov Ascunse Ierarhice; altele folosesc variante matematice ale Modelelor Markov Ascunse Ierarhice; iar altele utilizează abordări bazate pe reguli pentru codificarea unui set central de reguli sigure. URMA evaluează performanţele fiecărui sistem în condiţii propriu-zise de funcţionare şi le combină în mod optim. Există o doză de înţelegere greşită

în discuţiile publice privitoare la Watson, sub aspectul că oamenii de ştiinţă de la IBM care l-au creat se concentrează adeseori asupra URMA, managerul-expert produs de ei. Acest fapt conduce la comentarii de felul celui că Watson nu înţelege de fapt limba naturală deoarece este dificil să identifici în ce anume rezidă această înţelegere. Deşi scheletul URMA învaţă şi din propria experienţă, „înţelegerea” limbii de către Watson nu poate să se regăsească doar în URMĂ, ci este distribuită în toate numeroasele sale componente, inclusiv în modulele autoorganizatoare de limbă care folosesc metode similare Modelelor Markov Ascunse Ierarhice.

O parte separată a tehnologiei lui Watson face apel la estimarea încrederii, care ţine de URMA, pentru a determina felul cum să plaseze pariurile la *Jeopardy*! în vreme ce sistemul Watson este special optimizat pentru a juca acest joc, tehnologia lui de bază pentru căutarea în limbă şi cunoaştere poate fi adaptată cu uşurinţă în scopul îndeplinirii unor sarcini mai largi. Se poate crede că un areal de cunoştinţe mai puţin obişnuit, cum este domeniul medical, ar fi mult mai uşor de stăpânit decât cunoştinţele „obişnuite” necesare pentru a juca *Jeopardy*! De fapt, lucrurile stau exact invers: cunoştinţele profesionale tind să fie mai bine organizate şi structurate, deci mai puţin ambigue decât corespondentele lor comune, astfel încât folosirea acestor tehnici este deosebit de pretabilă la înţelegerea limbii naturale. După cum am menţionat, IBM lucrează actualmente cu Nuance pentru adaptarea tehnologiei lui Watson la medicină.

Conversaţia care are loc atunci când Watson joacă *Jeopardy*! este scurtă: se pune o întrebare, iar Watson dă răspunsul. (Din nou, din punct de vedere tehnic, vine cu o întrebare care să se potrivească răspunsului propus în concurs.) Nu se angajează într-o conversaţie care ar impune urmărirea tuturor afirmaţiilor anterioare ale tuturor participanţilor. (Siri face acest lucru, dar într-o măsură redusă: dacă îi ceri să-i transmită soţiei tale un

mesaj, îţi va cere să o identifici şi va reţine cine este aceasta pentru solicitările ulterioare.) Urmărirea întregii informaţii din cadrul unei conversaţii – sarcină care ar fi în mod evident impusă pentru a trece Testul Turing – e o cerinţă suplimentară semnificativă, dar nu cu mult mai dificilă decât ceea ce face deja Watson. La urma urmelor, Watson a citit sute de milioane de pagini de material, iar acesta a cuprins, în mod evident, şi numeroase relatări, astfel încât computerul este capabil să urmărească şi evenimente cu succesiuni complexe. Prin urmare, ar trebui să fie capabil să susţină propriile conversaţii şi să le ia în considerare în replicile ulterioare.

O altă limitare a jocului *Jeopardy*! rezidă în faptul că răspunsurile sunt în general scurte: de exemplu, nu se pun întrebări care să-i oblige pe concurenţi să numească cele cinci teme principale ale romanului *Poveste despre două oraşe*. În măsura în care poate găsi documente care să discute temele din acest roman, o versiune a lui Watson modificată corespunzător ar trebui să fie în stare să răspundă şi la această întrebare. Faptul de a avansa asemenea teme în mod independent, numai în urma lecturării cărţii şi nu prin copierea gândurilor altor gânditori (chiar şi fără cuvinte), reprezintă o altă chestiune. Această abordare este o altă sarcină de nivel superior pe care Watson se dovedeşte capabil s-o îndeplinească astăzi – e ceea ce aş numi sarcină de nivelul Testului Turing. (Acestea fiind spuse, voi scoate în evidenţă faptul că majoritatea oamenilor nu vin nici ei cu gânduri proprii şi originale, ci copiază ideile colegilor şi opiniile conducătorilor.) în orice caz, suntem în anul 2012, nu 2029, aşa că, deocamdată, nici nu mă aştept la o inteligenţă de nivelul Testului Turing. Dar, pe de altă parte, aş scoate în evidenţă faptul că evaluarea răspunsurilor la solicitări cum ar fi găsirea ideilor-cheie dintr-un roman nu este, ca atare, o sarcină simplă. Dacă eşti întrebat cine a semnat *Declaraţia de Independenţă*, poţi să evaluezi dacă răspunsul tău este adevărat sau nu. Validitatea

răspunsurilor la întrebări de nivel superior, cum ar fi descrierea temelor unei lucrări de creaţie, este mult mai greu de stabilit.

Demn de evidenţiat este şi faptul că, deşi aptitudinile lingvistice ale sale sunt, pe undeva, inferioare celor ale unei persoane umane cu o educaţie solidă, Watson s-a arătat capabil să-i înfrângă pe cei mai mari doi jucători de *Jeopardy*! din lume. Şi a reuşit să realizeze asta deoarece poate combina abilitatea lingvistică şi înţelegerea cunoştinţelor cu utilizarea perfectă şi de o remarcabilă acurateţe a memoriei pe care o posedă maşinile. Acesta este motivul pentru care le-am încredinţat maşinilor deja, într-o mare măsură, memoria noastră personală, socială şi istorică.

Cu toate că nu sunt pregătit să-mi devansez predicţia privitoare la un computer care să treacă Testul Turing înaintea anului 2029, progresele realizate în cazul unor sisteme ca Watson ar trebui să dea oricui o mare doză de încredere că apariţia inteligenţei artificiale de nivelul Testului Turing este aproape. Dacă ar trebui să se creeze o versiune a lui Watson optimizată pentru Testul Turing, probabil că s-ar ajunge foarte aproape de atingerea scopului.

Filosoful american John Searle (născut în anul 1932) a susţinut recent că Watson nu este capabil să gândească. Citându-şi propriul experiment mintal, denumit „Camera chinezească” (despre care voi discuta mai pe larg în Capitolul 11), el afirmă că Watson nu face decât să manipuleze simboluri şi că nu înţelege semnificaţia acestor simboluri. Practic, Searle nu îl descrie pe Watson cu acurateţe, deoarece înţelegerea limbajului de către acesta din urmă se bazează pe procese statistice ierarhice – nu pe manipularea simbolurilor. Singura modalitate de a aprecia drept exactă caracterizarea lui Searle ar fi aceea de a considera fiecare pas al proceselor de autoorganizare ale lui Watson ca fiind „manipulare a simbolurilor”. Dar, într-un asemenea caz, nici creierul omului nu ar putea fi

considerat capabil de judecată.

Este amuzant şi de râs faptul că observatorii îl critică pe Watson că nu face altceva decât o analiză statistică a limbii, în loc să proceseze „adevărata” înţelegere a limbii, care îi este proprie omului. Dar creierul uman efectuează exact o analiză statistică ierarhică atunci când rezolvă ipoteze multiple pe baza deducerii statistice (şi realmente la fiecare nivel al ierarhiei neocortexului). Atât Watson, cât şi creierul omului învaţă şi răspund pe baza unei abordări similare înţelegerii ierarhice. Din multe puncte de vedere, cunoştinţele lui Watson sunt mult mai cuprinzătoare decât ale omului; niciun om nu poate să pretindă că stăpâneşte toată Wikipedia, care reprezintă doar o parte din baza de cunoştinţe a lui Watson. Invers, astăzi omul poate stăpâni mai multe niveluri conceptuale decât Watson, dar în mod sigur deosebirea nu este definitivă.

Un sistem important care demonstrează forţa computaţiei aplicată cunoştinţelor organizate este Wolfram Alpha, un motor de răspunsuri (diferit de un motor de căutare) pus la punct de matematicianul şi omul de ştiinţă britanic dr. Wolfram (născut în anul 1959), împreună cu colegii săi de la Wolfram Research. De exemplu, dacă îl întrebi pe Wolfram Alpha (la

Wolfram Alpha.com) „Câte numere prime sunt într-un milion?”, acesta îţi va răspunde „78.498”. N-a căutat numărul acesta, ci l-a computat şi, după răspuns, îţi va da şi ecuaţiile pe care le-a folosit. Dacă ai încerca să ajungi la răspuns folosind un motor de căutare convenţional, acesta te-ar orienta către link-uri în care ai putea găsi algoritmii necesari. Pe urmă, ar trebui să introduci aceşti algoritmi într-un sistem cum ar fi *Mathematica*, pus la punct tot de dr. Wolfram, însă această abordare ţi-ar solicita în mod evident mult mai mult timp (şi capacitate de înţelegere) decât cel (e) necesar (e) ca să-l întrebi pur şi simplu pe Alpha.

Într-adevăr, Alpha constă în 15 milioane de linii de cod *Mathematica*. El poate computa răspunsul din

aproximativ 10 mii de miliarde de date care au fost atent alese de echipa de la Wolfram Research. Poţi să-i adresezi lui Alpha o gamă largă de întrebări efective, cum ar fi: „Ce ţară are cel mai mare PIB pe locuitor?” (Răspuns: „Monaco, 212.000 de dolari americani pe cap de locuitor”) sau „Ce vârstă are Stephen Wolfram?” (Răspuns: „52 de ani, 9 luni şi 2 zile” în momentul în care scriu aceste rânduri). După cum am menţionat, Alpha este folosit ca parte a lui Siri de la Apple; dacă îi adresezi lui Siri o întrebare referitoare la fapte reale, aceasta îi este transmisă mai departe lui Alpha, pentru a răspunde la ea. De asemenea, Alpha se ocupă de căutările solicitate pe motorul de căutare Bing al lui Microsoft.

Într-o postare recentă, dr. Wolfram a anunţat că Alpha dă actualmente răspunsuri corecte în 90 la sută dintre cazuri.17 De asemenea, a comunicat o reducere exponenţială a ratei eşecurilor, cu un timp de înjumătăţire de circa optsprezece luni. Este un sistem impresionant şi foloseşte metode create manual şi date verificate manual. E o explicaţie a motivului pentru care au fost create de fapt computerele. În timp ce noi descoperim şi compilăm metode ştiinţifice şi matematice, computerele se dovedesc a fi cu mult mai bune la implementarea lor decât inteligenţa umană neasistată. Cele mai multe metode ştiinţifice au fost codificate în Alpha laolaltă cu date actualizate permanent pe subiecte ce merg de la economie la fizică. În cursul unei conversaţii particulare pe care am purtat-o cu el, dr. Wolfram a estimat că metodele autoorganizatoare de felul celor folosite în cazul lui Watson ating o precizie de aproximativ 80 la sută atunci când funcţionează bine. Alpha, a menţionat dr. Wolfram, ajunge la o precizie de aproximativ 90 la sută. Desigur, în cazul ambelor date procentuale de mai înainte, intervine o autoselecţie, ceea ce înseamnă că utilizatorii (bunăoară, eu) au învăţat genul de întrebări la care este bun Alpha. Un proces similar se aplică şi metodelor autoorganizatoare. Prin urmare, optzeci la sută pare să fie

o estimare rezonabilă pentru precizia lui Watson la întrebările de la *Jeopardy*! iar ea a fost suficient de mare pentru ca Watson să-i înfrângă pe oamenii cei mai buni.

Opinia mea este că metodele autoorganizatoare de felul celor despre care am vorbit în cadrul Teoriei minţii bazate pe recunoaşterea modelelor sunt necesare pentru înţelegerea ierarhiilor complexe şi adeseori ambigue cu care ne confruntăm în cazul fenomenelor din lumea reală – inclusiv limba omului. O combinaţie ideală pentru un sistem inteligent robust ar presupune reunirea inteligenţei ierarhice susţinute de Teoria minţii bazate pe recunoaşterea modelelor (care afirmă că acesta este modul de funcţionare a creierului omului) cu codificarea precisă a cunoştinţelor şi datelor ştiinţifice. În esenţă, asta apropie omul de computer. Vom dezvolta ambii poli ai inteligenţei în anii care urmează. În ceea ce priveşte inteligenţa noastră biologică, deşi neocortexul prezintă o plasticitate semnificativă, arhitectura lui de bază este limitată de constrângerile fizice. Apariţia unui neocortex suplimentar în fruntea omului a fost o inovaţie importantă a evoluţiei, însă acum nu mai putem să ne mărim lobii frontali de o mie de ori, ba poate nici măcar cu 10 la sută. Cu alte cuvinte, nu putem realiza acest lucru în mod biologic, însă îl vom realiza prin mijloace tehnologice.

O strategie pentru crearea unei minţi în creierul nostru există miliarde de neuroni, dar ce sunt neuronii? Nişte simple celule. Creierul nu are niciun fel de cunoştinţe înainte de a se realiza conexiunile dintre neuroni. Tot ce ştim, tot ce suntem vine din felul în care sunt conectaţi neuronii.

— Tim Berners-Lee

Să folosim observaţiile pe care le-am discutat anterior pentru a începe să construim un creier. Vom porni de la realizarea unui recognizator de modele care să satisfacă toate cerinţele impuse. Pe urmă, vom face copii ale recognizatorului, în măsura în care pot să suporte memoria şi capacităţile noastre de computaţie. Orice

recognizator computează probabilitatea ca modelul lui să fie recunoscut. Procedând astfel, ia în considerare magnitudinea observată a fiecărui imput (în cadrul unui continuum adecvat) şi o compară cu parametrii de mărime şi de variabilitate a mărimii asociaţi fiecărui imput. Recognizatorul îşi excită axonul simulat, dacă probabilitatea computată depăşeşte un anumit prag. Acest prag şi parametrii care controlează computaţia probabilităţii modelului se numără printre elementele pe care le vom optimiza cu ajutorul unui algoritm genetic. Cum nu există cerinţa ca fiecare imput să fie activ pentru a fi recunoscut un anumit model, avem o recunoaştere autoasociativă (ceea ce înseamnă recunoaşterea unui model pe baza prezentării unei părţi a acelui model). De asemenea, permitem emiterea de semnale inhibitoare (semnale ce indică faptul că e mai puţin probabil să fie vorba de modelul respectiv). Recunoaşterea modelului determină generarea unui semnal activ pe axonul simulat al acestui recognizator de modele. Axonul este, la rândul lui, conectat cu unul sau mai mulţi recognizatori de modele de la nivelul conceptual superior. Toţi recognizatorii de modele conectaţi la următorul nivel conceptual superior acceptă acest model ca fiind unul dintre inputurile lor. Fiecare recognizator trimite semnale şi la recognizatorii de modele de nivel conceptual inferior, de câte ori recunoaşte un model, transmiţându-le că este „de aşteptat” restul modelului. Fiecare recognizator de modele are unul sau mai multe asemenea canale de imput „de aşteptat”. Când este primit, astfel, semnalul „de aşteptat”, pragul de recunoaştere pentru acest recognizator de modele e diminuat (e făcut mai accesibil).

Recognizatorii de modele sunt răspunzători pentru „legarea” la alţi recognizatori de modele, superiori sau inferiori, din ierarhia conceptuală. A se reţine că toate „legăturile” în cazul unei implementări software operează prin intermediul link-urilor virtuale (care, asemenea link-urilorweb, sunt în esenţă indicatoare de memorie) şi nu

prin legături propriu-zise. Acest sistem este, practic, mai flexibil decât cel din creierul biologic. În creierul uman, modelele noi trebuie să fie alocate unui recognizator de modele real, fizic, iar conexiunile noi trebuie realizate prin intermediul legăturii obişnuite axondendrită. De obicei, asta înseamnă că se ia o conexiune fizică existentă, adică elementul de care e nevoie, iar apoi se măresc extensiile necesare ale axonului şi dendritelor pentru realizarea conectării depline.

O altă tehnică folosită de creierul biologic al mamiferelor este să pornească de la un număr mare de conexiuni posibile şi apoi să taie selectiv conexiunile neuronale care nu sunt folosite. Dacă un neocortex biologic realocă recognizatori corticali de modele care au învăţat deja modele mai vechi pentru a asimila un material nou, conexiunile trebuie reconfigurate fizic. Din nou, aceste sarcini sunt mult mai simple în cazul unei implementări software. Pur şi simplu alocăm noi poziţii în memorie unui nou recognizator de modele şi folosim link-urile de memorie pentru conexiuni. Dacă neocortexul digital vrea să transfere resurse ale memoriei corticale de la un set de modele la altul, pur şi simplu returnează memoriei vechii recognizatori de modele şi apoi face noua alocare. Acest gen de „adunare a gunoiului” şi realocare a memoriereste o caracteristică-standard a arhitecturii multor sisteme software. În creierul nostru digital, am salvat şi amintirile vechi înainte să le descărcăm din neocortexul activ – măsură de prevedere pe care nu o putem lua în cazul creierului biologic.

Există o multitudine de tehnici matematice care pot fi folosite pentru implementarea acestei abordări privind recunoaşterea ierarhică autoorganizată a-modelelor. Metoda pe care aş folosi-o este aceea a Modelelor Markov Ascunse – şi asta din mai multe motive. Din perspectiva mea, am adunat câteva decenii de familiarizare cu această metodă, folosind-o la sistemele de pionierat de recunoaştere a vorbirii şi a limbii naturale, începând cu

anii 1980. Din perspectiva domeniului în ansamblul său, există pentru Modelele Markov Ascunse o experienţă mai cuprinzătoare decât pentru orice altă abordare a sarcinii de recunoaştere a modelelor. Modelele Markov Ascunse sunt folosite considerabil şi în înţelegerea limbii naturale. Numeroase sisteme de înţelegere a limbii naturale folosesc tehnici cel puţin matematic similare Modelelor Markov Ascunse Ierarhice.

Ase observa că nu toate sistemele bazate pe Modele Markov Ascunse sunt complet ierarhice. Unele permit existenţa a doar câteva niveluri ale ierarhiei – de exemplu, trecerea de la stările acustice la foneme şi apoi la cuvinte. Pentru construirea unui creier, vom dori să-i permitem sistemului nostru să creeze oricâte nivele ierarhice îi sunt necesare. De asemenea, majoritatea sistemelor bazate pe Modele Markov Ascunse nu sunt complet autoorganizate. Unele au conexiuni fixe, deşi aceste sisteme îşi taie efectiv multe dintre conexiunile de început, permiţându-le să ajungă la greutăţi de conexiune zero. Sistemele noastre din anii 1980 şi 1990 îşi tăiau automat conexiunile cu greutăţi aflate sub un anumit nivel şi, de asemenea, permiteau stabilirea unor noi conexiuni pentru a crea modele ale datelor de pregătire în condiţii mai bune şi pentru a învăţa din mers. O cerinţă-cheie, cred eu, este să i se permită sistemului să-şi creeze flexibil propriile topologii, pe baza modelelor la care este expus în cursul învăţării. Putem folosi tehnica matematică a programării liniare în scopul alocării optime a conexiunilor pentru noii recognizatori de modele.

Creierul nostru digital va acorda, de asemenea, o redundanţă substanţială fiecărui model, în special celor care apar frecvent. Aceasta permite o recunoaştere robustă a modelelor obişnuite şi este, de asemenea, una dintre metodeie-cheie pentru asigurarea unei recunoaşteri invariabile a diferitelor forme ale unui model. Pe de altă parte, vom avea nevoie de reguli cu privire la nivelul de redundanţă care poate fi admis, deoarece nu avem

interesul să folosim un volum excesiv de memorie pentru fiecare model obişnuit, de nivel redus.

Regulile referitoare la redundanţă, la pragurile de recunoaştere şi la efectul asupra pragului unei indicaţii de genul „Acest model este de aşteptat” sunt câteva dintre exemplele de parametri-cheie de ansamblu care afectează performanţa sistemului cu autoorganizare. Iniţial, aş stabili aceşti parametri pe baza intuiţiei mele, dar i-aş optimiza apoi folosind un algoritm genetic.

Un considerent foarte importanţii reprezintă educaţia creierului, indiferent dacă e vorba despre unul biologic sau de unul software. După cum am discutat anterior, un sistem ierarhic de recunoaştere a modelelor (digital sau biologic) va învăţa despre două – preferabil unui singur – niveluri ierarhice odată. Pentru a amorsa sistemul, aş porni de la reţele ierarhice pregătite anterior, care şi-au învăţat deja lecţiile de recunoaştere a vorbirii umane, a caracterelor tipărite şi a structurilor de limbă naturală. Un asemenea sistem ar fi capabil să citească documente în limbaj natural, dar nu ar putea să stăpânească la un moment dat decât un nivel conceptual aproximativ. Nivelurile asimilate anterior ar asigura o bază relativ stabilă pentru învăţarea următorului nivel. Sistemul poate citi acelaşi document în repetate rânduri, atingând noi niveluri conceptuale cu fiecare citire ulterioară, similar modului în care oamenii recitesc şi ajung la o înţelegere mai profundă a textelor. Pe web sunt disponibile miliarde de pagini de materiale. Wikipedia însăşi conţine în jur de patru milioane de articole în versiune engleză.

Aş oferi, de asemenea, un modul de gândire critică având proprietatea de a realiza o scanare continuă de fundal a tuturor modelelor existente, revizuindu-le compatibilitatea cu alte modele (idei) în acest neocortex software. Nu avem o asemenea facilitate în creierele noastre biologice, motiv pentru care oamenii pot să nutrească în mod egal gânduri complet inconsecvente. La identificarea unei idei inconsecvente, modulul digital

începe să caute o rezolvare, incluzând propriile structuri corticale şi întreaga literatură care îi stă la dispoziţie. O decizie poate să însemne, de pildă, determinarea faptului că una dintre ideile inconsecvente este pur şi simplu incorectă (dacă faptul e contraindicat de o preponderenţă a datelor aflate în conflict). Mai constructiv, ar găsi la un nivel conceptual superior o idee capabilă să rezolve aparenta contradicţie, asigurând o perspectivă ce explică fiecare idee. Sistemul ar adăuga această rezoluţie ca nou model şi ca link la ideile care au declanşat căutarea soluţiei. Acest modul critic de gândire ar sluji ca funcţie continuă de fundal. Ar fi foarte benefic dacă creierul uman ar face acelaşi lucru.

De asemenea, ar asigura un modul de identificare a întrebărilor deschise din fiecare disciplină. Ca altă funcţie continuă de fundal, ar căuta soluţii pentru respectivele întrebări în alte arii ale cunoaşterii. După cum am subliniat, la nivelul neocortexului, cunoştinţele constau în module profund înrădăcinate ale modelelor şi, prin urmare, sunt complet metaforice. Un model nu poate fi folosit decât pentru a asigura o soluţie, ori ca sursă de intuiţie pentru un domeniu aparent fără legătură.

Ca exemplu, aminteşte-ţi de metafora pe care am folosit-o în Capitolul 4, cea care lega mişcările aleatorii ale moleculelor într-un gaz de mişcările aleatorii ale schimbărilor evolutive. Moleculele dintr-un gaz se mişcă aleatoriu, fără niciun simţ aparent al direcţiei. În ciuda acestui fapt, dacă i se va acorda suficient timp, practic fiecare moleculă dintr-un gaz aflat într-un pahar de laborator va părăsi recipientul. Am observat că acest fapt asigură o perspectivă legată de o întrebare importantă referitoare la evoluţia inteligenţei. Asemenea moleculelor de gaz, schimbările evolutive se îndreaptă în toate părţile, aparent fără direcţie. Cu toate acestea, sesizăm o mişcare către o mai mare complexitate şi o mai mare inteligenţă, către realizarea supremă a evoluţiei unui neocortex capabil de gândire ierarhică. Aşadar, suntem capabili să ne

facem o idee despre modul în care un proces aparent lipsit de scop şi de direcţie poate avea un rezultat aparent plin de logică într-un domeniu (evoluţia biologică), îndreptându-ne atenţia către un alt domeniu (termodinamica).

Am menţionat mai devreme că ideea lui Charles Lyell privitoare la faptul că modificările minore provocate de apele curgătoare în corpul formaţiunilor stâncoase pot duce la apariţia unor văi mari l-a inspirat peste timp pe Charles Darwin să facă o observaţie similară în legătură cu modificările minime continue ale caracteristicilor organismelor în cadrul unei specii. Această căutare metaforică ar fi un alt proces continuu de fundal.

Ar trebui să asigurăm un mod de parcurgere a mai multor liste în mod simultan, pentru garantarea echivalentului gândului structurat. O listă poate însemna afirmarea constrângerilor pe care trebuie să le satisfacă soluţia la o problemă. Fiecare pas poate genera o căutare recursivă în ierarhia existentă de idei sau o căutare prin literatura disponibilă. Creierul uman pare a fi capabil să facă faţă doar unui număr de patru liste simultan (fără ajutorul unor instrumente precum computerul), dar nu există niciun motiv pentru care un neocortex artificial să aibă o asemenea limitare.

De asemenea, vom vrea să ne mărim creierul artificial cu genul de inteligenţă în care au excelat întotdeauna computerele, asta însemnând capacitatea de a stăpâni cu precizie vaste baze de date şi de a implementa algoritmi cunoscuţi în mod rapid şi eficient. Wolfram Alpha combină într-un fel unic numeroase metode ştiinţifice cunoscute şi le aplică unor date adunate atent. Acest tip de sistem va continua de asemenea să se îmbunătăţească, în virtutea observaţiei dr. Wolfram referitoare la un declin exponenţial al ratei erorilor.

În cele din urmă, creierul nostru are nevoie de scop. Scopul este exprimat ca o serie de scopuri. În cazul creierului biologic, scopurile ne sunt stabilite de centrii de

plăcere şi de frică pe care i-am moştenit de la creierul vechi. Aceste porniri primitive au fost generate iniţial de evoluţia biologică pentru asigurarea supravieţuirii speciilor, dar neocortexul ne-a permis să le sublimăm. Scopul lui Watson a fost să răspundă la întrebările de la *Jeopardy*! Un alt scop formulat simplu poate să fie trecerea Testului Turing. Pentru asta, creierul digital are nevoie de nararea umană a unei poveşti fictive proprii, astfel încât să poată pretinde că este uman din punct de vedere biologic. De asemenea, el trebuie să se „prostească" în mod considerabil, căci orice sistem care ar manifesta cunoştinţele lui Watson, să spunem, ar fi demascat repede ca fiind unul nonbiologic.

Fapt foarte interesant, am putea să-i dăm creierului nostru un scop mai ambiţios, ca de pildă acela de a contribui la o lume mai bună. Desigur, un scop de acest fel ridică o mulţime de întrebări: Mai bună pentru cine? Mai bună în ce fel? Pentru omul biologic? Pentru fiinţele conştiente? în acest caz, cine sau ce înseamnă conştient?

Pe măsură ce creierul nonbiologic devine tot atât de capabil precum cel biologic să efectueze schimbări în lume – pe bună dreptate, în cele din urmă mult mai capabil decât creierul biologic neperfecţionat –, va trebui să luăm în considerare educarea lui morală. Un bun punct de pornire l-ar putea reprezenta o veche idee din tradiţiile noastre religioase: Regula de Aur.

Capitolul 8
MINTEA CA UN COMPUTER

Având forma aproximativă a unei felii de pâine ţărănească din Franţa, creierul nostru este un aglomerat laborator de chimie, care debordează nonstop de conversaţii neuronale. Imaginează-ţi creierul: acei ghem lucios al existenţei, acel parlament de celule gri-şoarece, acea fabrică de vise, acel mic tiran plasat într-o bilă osoasă, acel talmeş-balmeş de neuroni care comandă toate actele, acel mic pretutindeni, acel capricios teren de distracţie, acea garderobă încreţită de euri înghesuite în

craniu, la fel ca prea multe haine într-o geantă pentru sală.

— Diane Ackerman

Creierul există pentru că distribuţia de resurse necesare supravieţuirii şi pericolele care ameninţă supravieţuirea variază în spaţiu şi timp.

— John NI. Allman

Geografia modernă a creierului are o calitate delicios antichizată – foarte asemănătoare unei hărţi cu lumea cunoscută încercuită de *terra incognila* în care bântuie monştri.

— David Bainbridge în matematică, nu trebuie să înţelegi lucrurile. Pur şi simplu te obişnuieşti cu ele.

— John von Neumann încă de la apariţia computerului, la mijlocul secolului douăzeci, a avut loc o dezbatere continuă nu numai pe tema extinderii finale a abilităţilor acestuia, ci şi dacă nu cumva însuşi cfeierui uman poate fi considerat o formă de computer. În ceea ce priveşte a doua temă, consensul a virat de la considerarea acestor două genuri de entităţi care procesează informaţii ca fiind esenţialmente identice la opinia că sunt fundamental diferite. Prin urmare, este creierul un computer?

Când au devenit pentru prima oară un subiect popular, în anii 1940, computerele au fost considerate imediat maşini gânditoare. Computerul ENIAC, anunţat în anul 1946, a fost descris în presă ca fiind un „creier gigantic”. Pe măsură ce computerele au devenit disponibile comercial în cursul următorului deceniu, publicitatea s-a referit în mod curent la ele ca la nişte creiere capabile de roluri cărora creierul obişnuit, biologic, nu le poate face faţă.

O reclamă din 1957 ce prezintă punctul de vedere popular, care considera computerul drept un creier gigantic

Programele de computer le-au permis maşinilor să se prezinte la înălţimea acestei aprecieri. „Rezolvitorul general de probleme”, creat în anul 1959 de Herbert A. Simon, J.C. Shaw şi Allen Newell la Universitatea Carnegie, a reuşit să facă demonstraţia unei teoreme pe care matematicienii Bertrand Russell (1872 – 1970) şi Alfred North Whitehead (1861 – 1947) nu reuşiseră să o demonstreze în celebra lor lucrare din anul 1913 *Principia Mathematica*. A devenit evident în cursul deceniilor care au urmat că aceste computere puteau să depăşească repede şi semnificativ capacităţile umane neasistate în cazul unor exerciţii intelectuale cum ar fi rezolvarea problemelor matematice, diagnosticarea bolilor şi jocul de şah, însă aveau dificultăţi în dirijarea unui robot să lege şireturile de la pantof sau în înţelegerea limbii obişnuite pe care putea să o priceapă un copil de cinci ani. Computerele abia încep acum să stăpânească acest gen de abilităţi. În mod ironic, evoluţia inteligenţei computerului a avut o direcţie opusă maturizării omului.

Problema dacă creierul uman şi computerul sunt, la un anumit nivel, echivalente rămâne controversată şi azi. În *Introducere* am menţionat că există milioane de link-uri de citate pe tema complexităţii creierului uman. În mod similar, o căutare Google pentru „Citate: Creierul nu este un computer” dă de asemenea milioane de link-uri. Din punctul meu de vedere, afirmaţiile pe această temă sunt echivalente cu a spune „Sucul concentrat de măr nu e măr”. Sub aspect tehnic, această afirmaţie este adevărată, dar poţi să faci suc concentrat de măr dintr-un măr. Poate şi mai aproape de subiect, asta este ca şi cum ai spune „Computerele nu sunt procesoare de cuvinte”. Este adevărat că atât computerul, cât şi procesorul de cuvinte există la nivele conceptuale diferite, dar un computer poate deveni procesor de cuvinte dacă rulează un software de procesare a acestora, nu şi invers. În mod similar, computerul poate deveni creier dacă rulează software de creier. Acesta este lucrul pe care vor să-l facă cercetătorii, inclusiv eu.

Şi atunci, întrebarea este dacă putem sau nu să găsim un algoritm care să transforme computerul într-o entitate echivalentă cu un creier uman. La urma urmelor, computerul poate rula orice algoritm pe care îl putem defini, datorită universalităţii lui înnăscute (supuse numai capacităţii lui). Creierul uman, pe de altă parte, funcţionează cu un set specific de algoritmi. Metodele lui sunt inteligente prin faptul că permit o semnificativă plasticitate şi restructurarea propriilor conexiuni pe baza experienţei sale, însă aceste funcţii pot fi imitate de un software.

Universalitatea computaţiei (ideea că un computer de uz general poate să implementeze orice fel de algoritm) şi forţa acestei idei au apărut în acelaşi timp cu primele maşini propriu-zise. Există patru concepte-cheie care stau la baza universalităţii, fezabilităţii şi aplicabilităţii computaţiei la gândirea umană. Ele merită să fie trecute în revistă aici, deoarece creierul însuşi le foloseşte. Primul îl

reprezintă abilitatea de comunicare, reţinere şi computare a informaţiei într-un mod sigur. În jurul anului 1940, dacă foloseai cuvântul „computer”, oamenii presupuneau că vorbeai despre un computer analogic, la care numerele erau reprezentate de diferite tensiuni electrice, iar componentele specializate puteau să realizeze operaţii matematice, cum ar fi adunarea şi înmulţirea. O mare limitare a computerelor analogice a constat în faptul că erau marcate de probleme de acurateţe. Numerele puteau fi reprezentate cu o precizie de numai o sutime, iar când tensiunile electrice care le reprezentau erau procesate prin creşterea numărului de operatori matematici, erorile se acumulau. Dacă voiai să realizezi ceva mai mult de o mână de operaţiit rezultatele deveneau atât de imprecise, încât îşi pierdeau sensul.

Oricine îşi mai aminteşte de vremurile în care muzica era înregistrată pe maşini cu bandă analogice îşi va aduce aminte şi de acest efect. La prima copie apărea o degradare sesizabilă, conţinând ceva mai mult zgomot decât originalul. (îţi aminteşti, dragă cititorule, că „zgomotul” reprezintă impreciziile aleatorii.) o copie a copiei conţinea şi mai mult zgomot, iar până la a zecea generaţie, copia însemna aproape complet numai zgomot. S-a presupus că aceeaşi problemă avea să marcheze şi apariţia lumii computerelor digitale. Putem înţelege asemenea preocupări dacă luăm în calcul comunicarea informaţiei digitale printr-un canal. Niciun canal nu este perfect şi fiecare prezintă o anumită rată de eroare. Să presupunem că avem un canal cu probabilitatea 0,9 de transmitere corectă a fiecărui bit. Dacă aş expedia un mesaj cu lungimea de un bit, probabilitatea transmiterii lui exacte pe acest canal ar fi de 0,9. Presupunem că transmitem doi biţi? Acum, acurateţea este de 0,92 = 0, 81. Dar dacă aş trimite un byte (opt biţi)? Am o şansă nu tocmai dreaptă (0,43, ca să fiu exact) de transmitere corectă a acestui byte. Probabilitatea expedierii precise a unui volum de informaţie de cinci byte este de aproximativ

1 la sută.

O soluţie evidentă pentru ocolirea problemei este ca acest canal să devină mai precis. Să presupunem că respectivul canal face doar o eroare la un milion de biţi. Dacă aş transmite un fişier constând în jumătate de milion de bytes (cam de proporţiile unui program modest sau ale unei baze de date), probabilitatea expedierii corecte ar fi mai mică de 2 la sută, în ciuda acurateţei foarte mari a canalului. Dat fiind faptul că, în cazul unui singur bit, eroarea poate invalida complet un program de computer şi alte tipuri de date digitale, situaţia nu este satisfăcătoare. Indiferent de acurateţea canalului, deoarece probabilitatea apariţiei unei erori într-o transmisie creşte rapid odată cu volumul mesajului, s-ar putea să pară că avem de-a face cu o barieră refractară.

Computerele analogice au abordat această problemă printr-o degradare graţioasă (ceea ce însemna că utilizatorii prezentau doar probleme în care se puteau tolera mici erori); însă dacă utilizatorii computerelor analogice se limitau la un set constrâns de calcule, computerele se dovedeau pe undeva utile. Computerele digitale, pe de altă parte, necesită o comunicare continuă, nu numai de la un computer la altul, ci şi la nivelul aceluiaşi computer. Aici există comunicare de la şi către memoriile sale, către şi de la unitatea de procesare. În unitatea centrală de procesare, există comunicare de la un registru la altul şi către şi dinspre unitatea aritmetică şi aşa mai departe. Chiar şi în cazul unităţii aritmetice există comunicare între un registru de biţi şi altul. Comunicarea este atotprezentă la toate nivelurile. Dacă ţinem seama de faptul că rata erorilor creşte rapid odată cu înmulţirea comunicaţiilor şi că o eroare de un singur bit poate distruge integritatea unui proces, computaţia digitală era condamnată – sau aşa părea la vremea respectivă.

Fapt remarcabil, aceasta a fost perspectiva obişnuită până când a apărut matematicianul american Claude Shannon (1916 – 2001) şi a demonstrat cum putem crea în

mod arbitrar o comunicare precisă folosind în acest scop chiar şi cele mai nesigure canale. Ce a afirmat Shannon în lucrarea lui fundamentală *A Mathematical Theory of Communication – O teorie matematică a comunicării*, publicată în *Bell Sistem Tehnical Journal* în lunile iulie şi octombrie 1948 şi în special în teorema lui privind codificarea pentru canale zgomotoase a fost că, dacă ai la dispoziţie un canal cu o rată de eroare oarecare (cu excepţia celor de exact 50 la sută pe bit, fapt ce ar însemna că ele ar transmite doar zgomot pur), poţi expedia un mesaj pentru care rata de eroare va fi cât de redusă vrei tu. Cu alte cuvinte, rata de eroare a transmisiunii poate fi de un bit din *n* biţi, unde *n* poate fi oricât de mare vrei tu să fie. Aşa, de exemplu, la extrem, dacă ai un canal care transmite în mod corect biţi de informaţie în numai 51 la sută din timp (respectiv, transmite bitul corect doar cu ceva mai multă frecvenţă decât bitul greşit), poţi totuşi să expediezi mesaje în care un singur bit dintr-un milion să fie incorect sau chiar unul singur dintr-o mie de miliarde sau dintr-un miliard de miliarde.

Cum e posibil aşa ceva? Răspunsul este: prin redundanţă. Acest lucru poate părea evident acum, dar nu a fost aşa şi la momentul respectiv. Ca exemplu simplu, dacă transmit fiecare bit de trei ori şi mă bazez pe votul majorităţii, sporesc substanţial fiabilitatea. Dacă nici aşa nu e suficient de bine, pur şi simplu măresc redundanţa până obţin fiabilitatea pe care mi-o doresc. Simpla repetare a informaţiei reprezintă modul cel mai simplu de asigurare a unor rate de acurateţe ridicate pe canale de acurateţe redusă, dar ea nu este şi cea mai eficientă abordare. Lucrarea lui Shannon, care a întemeiat domeniul teoriei informaţiei, prezenta şi metode optime sau coduri de detectare şi corectare a erorilor care pot asigura orice grad de acurateţe pe orice canal nealeatoriu.

Cititorii mai vârstnici îşi amintesc de modemurile telefonice, care transmiteau informaţia pe linii analogice cu nivel ridicat de zgomot. Aceste linii aveau un fond sonor

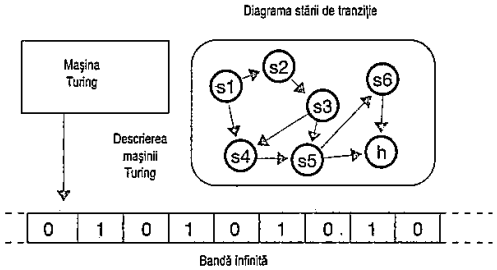
cu şuierături şi pocnete evidente auditiv, precum şi cu alte forme de distorsiuni, dar, în ciuda acestui fapt, reuşeau să transmită datele digitale cu rate foarte înalte de acurateţe, mulţumită teoremei canalului zgomotos a lui Shannon. Aceeaşi problemă şi aceeaşi soluţie există şi pentru memoria digitală. Te-ai întrebat vreodată, cititorule, cum reuşesc CD-urile, DVD-urile şi discurile cu programe să dea rezultate fiabile chiar şi după ce discul a fost scăpat pe jos sau zgâriat? Din nou ne putem gândi la Shannon.

Computaţia constă în trei elemente: comunicaţia – care, după cum am menţionat, este atotprezentă atât în interiorul, cât şi între computere –, memoria şi porţile logice (care îndeplinesc funcţiile aritmetice şi logice). Acurateţea porţilor logice poate fi de asemenea mărită, prin utilizarea similară a codurilor de detectare şi corectare a erorilor. Datorită teoremei şi teoriei lui Shannon putem noi acum să manevrăm volume arbitrar de mari sau complexe de date digitale şi algoritmi fără ca procesele să fie tulburate sau distruse de erori. Este important să evidenţiez faptul că creierul foloseşte şi el principiul lui Shannon, deşi evoluţia celui uman în mod clar îl antedatează pe Shannon! Majoritatea modelelor sau ideilor (şi o idee este tot un model), după cum am văzut, sunt stocate în creier cu un volum substanţial de redundanţă. Un motiv primar care explică redundanţa creierului este nefiabilitatea inerentă a circuitelor neuronale.

A doua idee importantă pe care se bazează epoca informaţiei este cea menţionată anterior: universalitatea computaţiei. În anul 1936, Alan Turing şi-a descris „maşina Turing”, care nu era propriu-zis o maşină, ci un alt experiment mintal. Computerul lui teoretic constă într-o bandă de memorie de lungime infinită cu un 1 sau o pe fiecare pătrat. Imputul maşinii este prezent pe această bandă, pe care maşina o poate citi pătrat cu pătrat. Maşina mai conţine şi un tabel de reguli – în esenţă, un program stocat – care constă în stări numărate. Fiecare regulă

specifică o acţiune dacă pătratul citit pe moment este un o şi o acţiune diferită dacă pătratul curent este 1. Posibilele acţiuni includ scrierea unui o sau 1 pe bandă, deplasarea benzii cu un pătrat la dreapta sau la stânga şi oprirea ei. Fiecare stare specifică, prin urmare, numărul stării următoare în care ar trebui să se afle maşina.

Imputul maşinii Turing este prezentat pe bandă. Programul rulează şi, când maşina se opreşte, şi-a încheiat algoritmul, iaroutputul procesului este lăsat pe bandă. A se observa că, deşi banda are, teoretic, o lungime infinită, orice program propriu-zis care nu se încadrează într-o buclă infinită va folosi doar o porţiune finită din bandă, aşadar, şi dacă ne limităm la o bandă finită, maşina tot va rezolva un set util de probleme.



O diagramă schematică a unei maşini Turing, cu capul care citeşte şi scrie pe bandă şi cu un program intern care constă în tranziţii de stare

Dacă maşina Turing pare simplă, asta se datorează faptului că acesta a fost obiectivul inventatorului ei. Turing a vrut ca maşina lui să fie cât mai simplă cu putinţă (dar nu mai simplă, ca să-l parafrazez pe Einstein). Turing şi Alonzo Church (1903 – 1995), fostul său profesor, au continuat şi au pus la punct Teza Church-Turing, care afirmă că, dacă o problemă ce poate fi pusă maşinii Turing nu poate fi rezolvată de aceasta, nu poate fi rezolvată de nicio altă maşină, potrivit legii naturale. Cu toate că maşina Turing are doar o mână de comenzi şi procesează doar câte un bit o dată, este capabilă să computeze orice pot computa alte computere. Un alt mod de a spune acest lucru este că orice maşină „completă Turing” (cu alte

cuvinte, care are capacităţi echivalente cu o maşină Turing) poate computa orice algoritm (orice procedură pe care o putem defini).

Interpretările „dure" ale Tezei Church-Turing propun o echivalenţă esenţială între ceea ce poate să gândească sau să ştie un om şi ceea ce poate computa o maşină. Ideea de bază este că creierul uman e supus şi el legii naturale şi, prin urmare, capacitatea lui de a procesa informaţia nu o poate depăşi pe cea a unei maşini (aşadar nici pe a maşinii Turing).

Îl putem credita pe bună dreptate pe Turing cu stabilirea fundamentului teoretic al computaţiei odată cu apariţia lucrării lui din anul 1936, dar este important de notat faptul că savantul a fost profund influenţat de o prelegere pe care matematicianul american de origine maghiară John von Neumann (1903 – 1957) a ţinut-o la Cambridge în anul 1935 pe tema conceptului său de program stocat, concept consfinţit de maşina Turing.1 La rândul lui, von Neumann a fost influenţat de lucrarea din anul 1936 a lui Turing, care a stabilit cu eleganţă principiile computaţiei şi a devenit o lectură obligatorie pentru colegii săi de la sfârşitul anilor 1930 şi începutul anilor 1940.2

În aceeaşi lucrare, Turing comunică o altă descoperire neaşteptată: aceea a problemelor insolvabile. Acestea sunt probleme bine definite, cu răspunsuri unice, care pot fi demonstrate că există, dar pe care putem dovedi că nu pot fi niciodată computate de vreo maşină Turing – cu alte cuvinte de nicio maşină, reversul a ceea ce a reprezentat o dogmă a secolului al nouăsprezecelea. Potrivit acesteia, problemele care pot fi definite vor fi în cele din urmă şi rezolvate. Turing a arătat că există tot atâtea probleme nerezolvabile câte probleme rezolvabile există. Matematicianul şi filosoful american de origine austriacă Kurt Godel a ajuns la o concluzie similară în a sa „Teoremă a incompletitudinii" din anul 1936. Rămânem astfel în situaţia perplexantă de a fi capabili să stabilim o

problemă, de a dovedi că există un răspuns unic, dar şi de a şti că, totuşi, răspunsul nu poate fi găsit vreodată.

Turing arătase că, în esenţa ei, computaţia se bazează pe un mecanism foarte simplu. Deoarece maşina Turing (şi orice computer, prin urmare) este capabilă să îşi întemeieze viitorul curs de acţiune pe rezultatele pe care le-a computat deja, ea e capabilă şi să ia decizii şi să realizeze modele de complexitate arbitrară a ierarhiilor de informaţii.

În anul 1939, Turing a proiectat un calculator electronic numit *Bombe*, care a ajutat la decodificarea mesajelor ce fuseseră criptate de maşina de codificat *Enigma* a naziştilor. În anul 1943, o echipă de ingineri influenţată de Turing a terminat ceea ce se poate demonstra că a fost primul computer, *Colossus*, care le-a permis Aliaţilor să continue decodificarea mesajelor criptate cu o versiune mai sofisticată a *Enigmei*. *Bombe* şi *Colossus* erau concepute pentru o sarcină unică şi nu puteau fi reprogramate pentru o alta. Dar se achitau de această sarcină într-un mod strălucit şi sunt creditate cu faptul că le-au permis Aliaţilor să depăşească avantajul de trei la unu de care se bucura aviaţia germană Luftwaffe în faţa Forţelor Aeriene Regale britanice şi să câştige cruciala Bătălie a Britaniei, dar şi să continue să anticipeze tactica naziştilor pe tot parcursul războiului;

Pe acest fundament a creat John Von Neumann arhitectura computerului modern, care reprezintă cea de-a treia idee majoră a noastră. Denumită „maşina Neumann”, ea a rămas practic structura centrală a oricărui computer din ultimii şaizeci şi şapte de ani, de la microcontrolul din maşina de spălat, la cele mai mari supercomputere. Într-o lucrare datată 30 iunie 1945 şi intitulată *Prima schiţă a unui raport despre EDVAC*, von Neumann a prezentat ideile care au dominat computaţia de atunci încoace.3 Modelul von Neumann include o unitate centrală de procesare, în care se realizează operaţiile aritmetice şi logice; o unitate de memorie, în care sunt stocate

programul şi datele; memoria de masă; un contor de program şi canalele de imputfoutput. Deşi această lucrare era destinată să fie un document intern al proiectului, a devenit Biblia creatorilor de computere. Nu poţi şti când se va întâmpla ca o memorie internă de rutină să pună capăt revoluţionării lumii.

Nu s-a intenţionat ca maşina Turing să fie practică. Teoremele lui Turing s-au ocupat nu de eficienţa rezolvării problemelor în examinarea gamei de probleme care puteau fi rezolvate, teoretic, de computaţie. Scopul lui von Neumann, pe de altă parte, a fost acela de a crea un concept fezabil de maşină de computat. Modelul lui înlocuieşte computaţiile bit-cubit ale lui Turing cu versiunea bit-multiple (în general, multipli de opt biţi). Banda de memorie a lui Turing este secvenţială, astfel încât programele pentru maşina Turing consumă excesiv de mult timp pentru mutarea benzii înainte şi înapoi în scopul stocării şi recuperării rezultatelor intermediare. Dimpotrivă, memoria lui von Neumann are acces aleatoriu, astfel încât orice element de date poate fi recuperat imediat.

Una dintre ideile-cheie ale lui von Neumann a fost programul stocat, pe care îl introdusese cu un deceniu înainte: plasarea programului în acelaşi tip de memorie cu acces aleatoriu ca şi datele (şi adeseori în acelaşi bloc de memorie). Acest fapt permite ca un computer să fie reprogramat atât pentru diferite funcţii, cât şi pentru codul de automodificare (dacă stocarea programului este inscriptibilă), operaţie ce facilitează o formă eficientă de recursivitate. Până atunci, practic toate computerele, inclusiv *Colossus*, fuseseră construite pentru o funcţie specifică. Programul stocat oferă unui computer posibilitatea de a fi cu adevărat universal, adeverind prin asta viziunea lui Turing privitoare la universalitatea computaţiei.

Un alt aspect-cheie al maşinii von Neumann este acela că fiecare instrucţiune include un cod ce specifică

operaţia aritmetică sau logică pe care maşina urmează să o realizeze şi adresa unui operant din memorie.

Conceptul lui von Neumann cu privire la modul cum trebuie proiectat un computer a fost lansat odată cu publicarea proiectului EDVAC, realizat împreună cu colaboratorii J. Presper Eckert şi John Mauchly. EDVAC nu a funcţionat propriu-zis până în anul 1951, moment în care existau şi alte computere cu programe stocate, precum Manchester Small-Scale Experimental Machine, ENIAC, EDSAC şi BINAC, toate profund influenţate de lucrarea lui von Neumann şi avându-i ca proiectanţi pe Eckert şi Mauchly. Von Neumann a contribuit direct la conceperea mai multor maşini de acest fel, inclusiv a unei versiuni ulterioare a lui ENIAC, care includea un program stocat.

Au existat câţiva precursori ai arhitecturii lui von Neumann, deşi, cu o excepţie surprinzătoare, nici măcar unul nu a fost o maşină von Neumann veritabilă. În anul 1944, Howard Aiken a lansat Mark I, care avea un element de programabilitate, dar nu folosea un program stocat. Citea instrucţiuni de pe o bandă de hârtie perforată şi apoi executa imediat comanda, însă îi lipsea o instrucţiune pe ramură condiţională.

În anul 1941, omul de ştiinţă german Konrad Zuse (1910 – 1995) a creat computerul Z-3. Şi acesta putea să-şi citească programul de pe o bandă (în cazul de faţă, codificat pe film), dar nu avea nici el instrucţiuni pe ramură condiţională. Fapt interesant, Zuse s-a bucurat de ajutor din partea Institutului German pentru Cercetări Aeronautice, care a folosit acest dispozitiv pentru Studierea vibraţiilor aripilor de avion, dar propunerea făcută de el guvernului nazist, de finanţare a înlocuirii releelor sale cu tuburi vidate, a fost respinsă. Naziştii au considerat că acest domeniu al computaţiei nu era „important pentru război”. Această perspectivă contribuie foarte mult, din punctul meu de vedere, la explicarea rezultatului războiului.

De fapt, există un precursor veritabil al conceptului

lui von Neumann şi acesta vine de dinainte de el cu un secol! Motorul Analitic al matematicianului şi inventatorului englez Charles Babbage (1791 – 1871), care a fost descris pentru prima oară în anul 1837, încorpora ideile lui von Neumann şi avea un program stocat pe cartoane perforate, împrumutate de la războiul de ţesut Jacquard.4 Memoria cu acces aleatoriu a Motorului includea 1.000 de cuvinte de 50 de cifre zecimale fiecare (echivalentul a 21 de kilobiţi). Fiecare instrucţiune includea un cod de operare şi un număr operativ, exact ca limbajele pentru maşinile moderne. Includea şi ramificarea condiţională şi buclele, astfel încât era o maşină von Neumann veritabilă. Motorul Analitic era exclusiv mecanic şi se pare că depăşea abilităţile de concepţie şi organizare ale lui Babbage, care a construit componente ale acestei maşini, dar ea nu a funcţionat niciodată. Este neclar dacă pionierii computerului din secolul XX, inclusiv von Neumann, erau la curent cu cercetările lui Babbage.

Computerul lui Babbage s-a concretizat într-adevăr în crearea domeniului numit programare software. Scriitoarea engleză Ada Byron (1815 – 1852), contesă de Lovelace şi singurul copil legitim al poetului Lord Byron, a fost primul programator de computer din istorie. Ea a Scris pentru Motorul Analitic programe pe care trebuia să le corecteze în minte (de vreme ce computerul nu a funcţionat niciodată), practică foarte bine cunoscută inginerilor software din zilele noastre sub denumirea de „verificare la masă”. Lovelace a tradus un articol al matematicianului italian Luigi Menabrea pe tema Motorului Analitic şi a adăugat textului note proprii cuprinzătoare, scriind că „Motorul Analitic ţese modele algebrice exact la fel cum războiul de ţesut Jacquard ţese flori şi frunze”. Ada Byron a mers şi mai departe şi a făcut probabil primele speculaţii cu privire la fezabilitatea inteligenţei artificiale, dar a concluzionat că Motorul Analitic nu are „niciun fel de pretenţii de a da naştere la ceva”.

Concepţia lui Babbage este absolut miraculoasă dacă e să ţii seama de epoca în care a trăit şi a lucrat el. În orice caz, până la mijlocul secolului XX, ideile lui se pierduseră în negura timpului (cu toate că au fost redescoperite ulterior). Von Neumann are meritul de a fi cel care a conceptualizat şi articulat principiile-cheie ale computerului aşa cum îl cunoaştem noi astăzi, iar lumea recunoaşte acest fapt continuând să se refere la maşina von Neumann ca la principalul model de computaţie. Reţine, totuşi, cititorule, că maşina von Neumann comunică în permanenţă date între diferitele sale unităţi şi în cadrul acestor unităţi, astfel încât nu putea fi construită fără teoremele lui Shannon şi metodele pe care le-a conceput acesta pentru transmiterea şi stocarea informaţiei digitale sigure.

Iar asta ne conduce la Cea de-a patra idee importantă, care ne invită să trecem dincolo de concluzia Adei Byron că un computer nu poate gândi în mod creator şi nu poate găsi algoritmi-cheie folosiţi de creier şi să purcedem la transformarea computerului într-un creier. Alan Turing a lansat acest obiectiv în lucrarea sa din anul 1950 *Computing Machinery and Intelligence – Maşini şi inteligenţă computabile*, care include actualmente celebrul său Test Turing pentru stabilirea certă a faptului că o inteligenţă artificială a atins nivelul uman de inteligenţă.

În anul 1956, von Neumann a început să pregătească o serie de prelegeri destinate prestigioasei serii Silliman de la Universitatea Yale. Din cauza ravagiilor cancerului, savantul nu a mai susţinut aceste prelegeri şi nici nu a terminat manuscrisul de pe care urmau să fie citite ele. Acest document neterminat rămâne totuşi un precursor strălucit şi profetic al proiectului pe care eu îl consider a fi cel mai îndrăzneţ şi mai important al omenirii. A fost publicat postum, sub titlul *The Computer and the Brain – Computerul şi creierul*, în anul 1958. Este cât se poate de potrivit ca ultima lucrare a unuia dintre cei mai străluciţi matematicieni ai secolului XX, unul dintre pionierii erei

computerelor, să fie întruchiparea inteligenţei. Acest proiect a reprezentat cea mai serioasă cercetare timpurie a creierului uman din perspectiva matematicianului şi a specialistului în computere. Înaintea lui von Neumann, domeniile ştiinţei computerelor şi neuroştiinţei erau două insule fără vreo punte între ele.

Von Neumann îşi începe discuţia scoţând în evidenţă asemănările şi deosebirile dintre computer şi creierul uman. Dat fiind momentul în care a fost elaborat, acest manuscris este remarcabil de exact. Von Neumann a notat faptul că outputul neuronilor este digital – axonul fie emite, fie nu emite. Acest lucru era departe de a fi evident la vremea respectivă. Însă procesarea la nivelul dendritelor care duc la neuron şi în corpul propriu-zis al neuronului s-a dovedit a fi analogică, iar savantul şi-a prezentat calculele sub forma însumării greutăţii inputurilor cu un prag. Acest model de funcţionare a neuronului a condus la domeniul conectivităţii, care a construit sisteme bazate pe respectivul model atât ca hardware, cât şi ca software. (După cum am arătat în capitolul anterior, primul asemenea sistem a fost creat de Frank Rosenbiatt ca program software pe un computer IBM 704 la Universitatea Cornell, în anul 1957, imediat după ce au devenit accesibile prelegerile lui von Neumann.) Actualmente avem modele mai sofisticate ale felului în care neuronii combină inputurile, dar ideea fundamentală a procesării analogice a inputului dendritelor folosind concentraţii ale neurotransmiţătorilora rămas validă.

Von Neumann a aplicat conceptul universalităţii cortiputaţiei pentru a concluziona că, deşi arhitectura şi cărămizile fundamentale par să fie radical diferite în raportul creier-computer, putem totuşi concluziona că o maşină von Neumann poate simula procesarea la nivelul creierului.

Reciproca nu este însă valabilă, deoarece creierul nu e o maşină von Neumann şi nu are un program stocat ca

atare (deşi în mintea noastră putem simula o maşină Turing foarte simplă). Algoritmul şi metodele sunt implicite în structura acesteia. Von Neumann concluzionează corect că neuronii pot învăţa din inputuri modele, despre care am stabilit deja că sunt codificate parţial în tăria dendritelor. Ceea ce nu se ştia pe vremea lui von Neumann era faptul că învăţarea are loc şi prin crearea şi distrugerea conexiunilor dintre neuroni.

Von Neumann sublinia, anticipând, că viteza procesării neuronale este extrem de redusă, de ordinul sutelor de calcule pe secundă, dar că creierul compensează acest neajuns prin procesarea masiv paralelă – o altă intuiţie, deloc evidentă, dar cheie. Von Neumann a susţinut că toţi cei 1010 neuroni ai creierului (o apreciere destul de precisă; estimările actuale sunt cuprinse între 1010 şi 1011) procesează în acelaşi timp cu ceilalţi. De fapt, toate conexiunile (cu o medie de 103 până la 104 conexiuni pentru un neuron) computează simultan.

Von Neumann şi descrierile pe care le face el procesării neuronale sunt remarcabile, dată fiind starea de început a neuroştiinţei la momentul respectiv. Un aspect al lucrării sale cu care nu sunt de acord îl reprezintă evaluarea făcută de el în legătură cu capacitatea memoriei creierului. Savantul presupune că creierul îşi aminteşte fiecare imput pentru tot restul vieţii. Von Neumann ia în calcul o durată medie de viaţă de 6Q de ani sau de aproximativ 2 x 109 secunde. La aproximativ 14 inputuri pentru fiecare neuron în fiecare secundă (ceea ce înseamnă de fapt cu cel puţin trei ordine de mărime mai puţin) şi la 1010 neuroni, el ajunge la o estimare de circa 1020 de biţi pentru întreaga capacitate a memoriei creierului. Realitatea, după cum am subliniat anterior, este că ne amintim numai un eşantion foarte redus din gândurile şi experienţele noastre şi că amintirile nu sunt stocate ca modele în biţi la un nivel inferior (cum ar fi o imagine video), ci mai curând ca succesiuni de modele de nivel superior.

Descriind fiecare mecanism al creierului, von Neumann arată că un computer modern ar putea face acelaşi lucru în ciuda deosebirilor aparente. Mecanismele analogice ale creierului pot fi simulate prin cele digitale, deoarece computaţia digitală poate emula valori analogice la orice grad de precizie (iar precizia informaţiei analogice din creier este relativ redusă).

Paralelismul masiv al creierului poate fi simulat şi el, dat fiind avantajul semnificativ al vitezei computerelor în computaţia în serie (un avantaj care s-a extins foarte mult de-a lungul timpului), în plus, putem folosi procesarea în paralel la computere prin utilizarea maşinilor paralele von Neumann – exact cum funcţionează computerele din zilele noastre.

Von Neumann concluzionează că metodele creierului nu pot implica algoritmi secvenţiali lungi, dacă se ia în considerare rapiditatea cu care pot oamenii să ia decizii, combinată cu viteza foarte redusă de computare a neuronilor. Când jucătorul de la baza a treia are o minge în joc şi decide să o arunce spre prima bază şi nu spre a doua, ia această decizie într-o fracţiune de secundă, adică un timp suficient numai pentru ca fiecare neuron să treacă prin câteva cicluri. Von Neumann concluzionează corect că puterile remarcabile ale creierului sunt conferite de faptul că toate cele 100 de miliarde de neuroni sunt capabile să proceseze informaţia simultan. După cum am subliniat, cortexul vizual face judecăţi vizuale sofisticate în numai trei sau patru cicluri neuronale.

Creierul are o plasticitate considerabilă, care ne permite să învăţăm, însă computerul are o plasticitate şi mai mare, care poate să-i restructureze complet metodele, prin modificarea software-ului său. Prin urmare, sub acest aspect, computerul va putea să imite creierul, însă reciproca nu este valabilă.

Când von Neumann a comparat capacitatea organizării masiv paralele a creierului cu cea a computerelor vremii (cele câteva), a devenit clar că

creierul avea o capacitate mult mai mare de memorie şi viteză. A fost construit deja primul supercomputer care a îndeplinit specificaţiile corespunzătoare unora dintre cele mai conservatoare estimări ale vitezei necesare pentru simularea funcţională a creierului, uman (în jur de 1016 operaţii pe secundă).5 (Personal, estimez că acest nivel de computare va costa 1.000 de dolari până la începutul anilor ’2020.) în privinţa memoriei, suntem chiar şi mai aproape. Deşi şi-a redactat acest remarcabil manuscris devreme în cadrul istoriei computerului, von Neumann a avut totuşi încredere că atât hardware-ul, cât şi software-ul inteligenţei umane îşi vor afla în cele din urmă locul, aceasta fiind şi motivaţia lui pentru pregătirea prelegerilor despre care am pomenit.

Von Neumann a fost cât se poate de conştient de ritmul tot mai accelerat al progresului şi de implicaţiile profunde ale acestuia în viitorul omenirii. La un an după moartea sa, în anul 1957, confratele matematician Stanulam l-a citat că ar fi spus, la începutul anilor 1950, că „progresul din ce în ce mai accelerat al tehnologiei şi schimbările modului de viaţă al omului dau aparenţa apropierii unei singularităţi esenţiale în istoria rasei, dincolo de care treburile omului, aşa cum le cunoaştem, nu vor mai putea continua”. Aceasta este prima utilizare cunoscută a cuvântului „singularitate” în contextul istoriei tehnologice umane.

Ceea ce a intuit în mod fundamental von Neumann a fost faptul că există o echivalenţă esenţială între computer şi creier. A se observa că inteligenţa emoţională a omului biologic este parte a inteligenţei sale generale. Dacă intuiţia lui von Neumann este corectă şi dacă se acceptă ideea mea că o entitate nonbiologică în stare să recreeze inteligenţa (emoţională şi de alte genuri) a unui om biologic este conştientă (a se vedea capitolul următor), atunci ar trebui să se ajungă la concluzia că există o echivalenţă esenţială între computer-cu software-ul adecvat – şi minte (mintea conştientă). Prin urmare, von

Neumann a avut dreptate?

Majoritatea computerelor din zilele noastre sunt exclusiv digitale, în vreme ce creierul omului combină metodele digitală şi analogică. Dar metodele analogice sunt uşor şi rutinier recreate de cele digitale la orice nivel de acurateţe. Specialistul american în ştiinţa computerelor Carver Mead. (născut în 1934) a arătat că putem să imităm în siliciu în mod direct metodele analogice ale creierului, fapt demonstrat cu ceea ce el numea cipuri „neuromorfice”.6 Mead a arătat cum este posibil ca această metodă să fie de mii de ori mai eficientă decât metodele digitale care le imită pe cele analogice. Cu cât codificăm algoritmul masiv recurent al neocortexului, cu atât devine mai logică folosirea abordării lui Mead. Grupul de Computare Cognitivă de la IBM, condus de Dharmendra Modha, a lansat cipuri care imită neuronii şi conexiunile lor, având inclusiv capacitatea de a stabili noi conexiuni.7 Denumit SYNAPSE, unul dintre cipuri asigură o simulare directă a unui număr de 256 de neuroni cu aproximativ un sfert de milion de conexiuni sinaptice. Scopul proiectului este crearea unui neocortex simulat cu 10 miliarde de neuroni şi 100 de mii de miliarde de conexiuni – aproape cât un creier uman –, care foloseşte doar un kilowatt de putere.

Aşa cum a spus von Neumann în urmă cu peste o jumătate de secol, creierul este extrem de lent, dar masiv paralel. Circuitele digitale din zilele noastre sunt de cel puţin 10 milioane de ori mai rapide decât comutatoarele electrochimice ale creierului. Reciproc, toţi cei 300 de recognizatori de modele din neocortexul creierului procesează simultan şi întregul număr de o sută de mii de miliarde de conexiuni interneuronale computează concomitent.

Însă aspectul-cheie în ceea ce priveşte asigurarea hardware-ului necesar pentru realizarea cu succes a modelului unui creier uman îl reprezintă memoria de ansamblu şi capacitatea de calcul impusă. Noi nu trebuie

să copiem direct arhitectura creierului, aceasta fiind o abordare foarte ineficientă şi inflexibilă.

Hai să estimăm care ar fi aceste cerinţe hardware. Multe proiecte au încercat să irnite tipul ierarhic de învăţare şi de recunoaştere a modelelor care se realizează la nivelul neocortexului, inclusiv cercetările mele cu Modele Markov Ierarhice Ascunse. O estimare conservatoare extasă din propria mea experienţă spune că imitarea unui ciclu într-un singur recognizator de modele din neocortexul biologic al creierului ar avea nevoie de un număr de circa 3.000 de calcule. Majoritatea simulărilor se realizează la o jumătate din această estimare. Când creierul funcţionează cu circa 102 (100) de cicli pe secundă, ajungem la 3 * 105 (300.000) de calcule pe secundă pe recognizator de modele. Folosind estimarea mea de 3 x 10° (300 de milioane) de recognizatori de modele, obţinem în jur de 1014 (100 de mii de miliarde) de calcule pe secundă, o cifră care corespunde estimării mele din cartea *The Singularity Is Near-Singularitatea este aproape*. În această carte, am anticipat că simularea funcţională a creierului ar impune între 1014 şi 1016 calcule pe secundă (cos) şi am folosit 1016 calcule pentru a fi conservator. Estimarea lui Hans Moravec, specialist în inteligenţă artificială, bazată pe extrapolarea cerinţelor computaţionale ale procesării vizuale timpurii (iniţiale) la nivelul întregului creier este de 1014 cos, cifră ce corespunde estimării mele.

Maşinile obişnuite de birou pot să ajungă la 1010 cos, deşi acest nivel de performanţă poate fi amplificat semnificativ prin folosirea resurselor din nor. Cel mai rapid supercomputer, K Computer din Japonia, a ajuns deja la 1018 cos.8 Dat fiind faptul că algoritmul neocortexului este masiv repetat, abordarea folosirii de cipuri neuromorfice, aşa cum este IBM SYNAPSE, menţionat mai înainte, este de asemenea promiţătoare.

În ceea ce priveşte necesarul de memorie, ne trebuie în jur de 30 de biţi (aproximativ patru byte) pentru ca o

conexiune să se lege de unul din totalul de 300 de milioane de recognizatori de modele. Dacă estimăm o medie de opt inputuri pentru fiecare recognizator, ajungem la 32 de byte pentru unul singur. Dacă adăugăm o greutate de un byte pentru fiecare imput, ajungem la 40 de byte, Mai punem 32 de byte pentru conexiunile descendente şi ajungem la 72 de byte. Observă, cititorule, că numărul care cuantifică ramificarea în jos şi în sus va fi întotdeauna mult mai mare de opt, cu toate că aceşti arbori foarte ramificaţi sunt împărţiţi între ei de numeroşi recognizatori. De exemplu, pot să existe sute de recognizatori implicaţi în recunoaşterea literei „p”. Aceştia transmit semnale câtorva mii de asemenea recognizatori de la următorul nivel, care se ocupă de cuvinte şi expresii ce conţin „p”. Pe de altă parte, recognizatorii de „p” nu repetă arborele de conexiuni care deserveşte toate cuvintele şi expresiile ce conţin „p” – toţi folosesc în comun un asemenea arbore de conexiuni. Aceeaşi observaţie este valabilă şi pentru conexiunile descendente: recognizatorul răspunzător pentru cuvântul „APPLE0 le va spune miilor de recognizatori ai lui „E” de la un nivel inferior că este de aşteptat un „E”, dacă au fost identificate deja „A”, „P”, „P° şi „L”. Acest arbore de conexiuni nu se repetă pentru fiecare recognizator de cuvânt sau de expresie care vrea să informeze următorul nivel inferior că este de aşteptat un „E°. Iarăşi, este folosit în comun. Pentru acest motiv, o estimare de ansamblu de opt ascendent şi opt descendent – în medie, desigur – pentru un recognizator de model este ceva rezonabil. Chiar dacă mărim această estimare, ea nu modifică semnificativ ordinul de mărime al estimării rezultante.

Cu 3 x 108 (300 de milioane) de recognizatori de modele la 72 de byte fiecare, obţinem un necesar total de memorie de aproximativ 2 * 1010 (20 de miliarde) de byte. Este o cifră cât se poate de modestă, pe care computerele uzuale din zilele noastre o pot depăşi.

Aceste estimări sunt făcute doar ca să ofere

aproximări ale ordinului de mărime cerut. Dat fiind faptul că circuitele digitale sunt în mod inerent de circa 10 milioane de ori mai rapide decât circuitele neocorticale biologice, nu trebuie să egalăm paralelismul creierului uman – o procesare paralelă modestă (comparativ cu paralelismul de miliarde de ori mai mare al creierului uman) va fi suficientă. Se poate vedea că necesităţile computaţionale ajung astfel la îndemână. Recablarea permanentă a creierului – dendritele îşi creează noi sinapse continuu – poate fi de asemenea imitată de software cu ajutorul link-urilor, un sistem mult mai flexibil decât cel care implică plasticitatea creierului, impresionantă dar limitată.

Redundanţa folosită de creier pentru atingerea unor rezultate invariante solide poate fi imitată în mod cert în emulaţiile software. Suportul matematic al optimizării acestor tipuri de sisteme de învăţare ierarhice autoorganizate este bine înţeleasă. Organizarea creierului e departe de a fi optimă. Desigur, nici n-a trebuit să fie – a trebuit să fie doar suficient de bună pentru atingerea pragului de la care creierul să poată fi capabil să creeze instrumente care să-i compenseze propriile limitări.

O altă restricţie a neocortexului uman spune că nu există vreun proces care să elimine sau măcar să treacă în revistă ideile contradictorii, fapt ce explică de ce gândirea umană este adeseori masiv inconsecventă. Mecanismul care se ocupă de această aşa-numită gândire critică e slab, iar abilitatea nu-i este exercitată nici pe departe atât de des cât ar trebui. Într-un neocortex bazat pe software, putem include un proces care să dezvăluie contradicţiile, pentru a le verifica pe mai departe.

Este important să se observe că alcătuirea unei întregi regiuni a creierului este mai simplă decât compunerea unui singur neuron. Aşa cum am discutat anterior, modelele devin adeseori mai simple la un nivel superior – să ne gândim la analogia cu un computer. Trebuie neapărat să înţelegem fizica semiconductorilor

pentru a crea modelul unui tranzistor, iar ecuaţiile aflate la baza unui singur tranzistor real sunt complexe. Un circuit digital care multiplică două numere are nevoie de sute. Cu toate acestea, putem crea un model al acestui circuit de multiplicare foarte simplu, cu una sau două formule. Pentru un întreg computer, cu miliarde de tranzistori, se poate face un model prin intermediul setului de instrucţiuni şi ţal descrierii registrelor acestuia, descriere ce necesită câteva foi pentru text şi formule. Programele software pentru un sistem de operare, compilatoarele de limbaj şi asamblatoarele sunt complexe la un nivel rezonabil, însă realizarea modelului unui anumit program – de exemplu, un program de recunoaştere a vorbirii bazat pe Modele Markov Ascunse Ierarhice – poate fi la fel de bine descris pe numai câteva pagini cu ecuaţii. Nicăieri într-o asemenea descriere nu s-ar putea găsi detalii privitoare la fizica semiconductorilor sau chiar la arhitectura computerului.

O observaţie similară este adevărată şi în cazul creierului. Un recognizator neocortical de modele care detectează o anumită caracteristică vizuală invariabilă (de pildă, un chip), care realizează o filtrare de *bandpass* (restricţionarea inputului la o gamă de frecvenţe specifice) a sunetului sau care evaluează proximitatea temporală a doua evenimente poate fi descris cu mult mai puţine detalii specifice decât relaţiile fizice şi chimice propriu-zise care controlează neurotransmiţătorii, canalele de ioni şi alte variabile sinaptice şi dendritice implicate în procesele neuronale. Deşi această complexitate trebuie să fie atent cântărită înaintea trecerii la următorul nivel conceptual superior, mare parte din ea poate fi simplificată, pe măsură ce sunt dezlegate principiile de funcţionare a creierului.

Capitolul 9
EXPERIMENTE MINTALE DESPRE MINTE

Minţile sunt pur şi simplu ceea ce fac creierele.
— Marvin Minsky, Societatea Minţii

Când se vor construi maşini inteligente, nu ar trebui

să fim surprinşi să constatăm că acestea sunt la fel de confuze şi încăpăţânate ca oamenii în convingerile lor referitoare la subiectul minţii, al conştientei, al liberului arbitru şi al altora asemenea.

— Marvin Minsky, Societatea Minţii

Cine este conştient?

Adevărata istorie a conştientei începe cu prima minciună a cuiva.

— Iosif Brodsky

Suferinţa este unica origine a conştiinţei.

— Feodor Dostoievski*, însemnări din subterană*

Există un soi de plantă care, cu ajutorul florilor sale, consumă hrană organică: atunci când pe floare se aşâză o muscă, petalele i se închid şi ţin musca strâns până când planta o absoarbe în organismul ei; dar floarea nu se închide decât la ce e bun de mâncat; unei picături de ploaie sau unui beţişor nu-i acordă vreo atenţie. Ciudat ca un lucru atât de lipsit de conştiinţă să fie atât de atent la propriul interes. Dacă asta înseamnă inconştienţă, care e folosul conştienţei?

— Samuel Butler, 1871

Am examinat creierul ca entitate capabilă de anumite niveluri de a duce lucrurile la bun sfârşit. Dar această perspectivă, în esenţă, ne lasă în afara tabloului. Părem să trăim în creierul nostru. Avem vieţi subiective. Cum se relaţionează concepţia noastră obiectivă asupra creierului, pe care am discutat-o până acum, cu sentimentele noastre, cu sentimentul că suntem persoana care trăieşte aceste experienţe?

Filosoful britanic Colin McGinn (născut în anul 1950) scrie că discutarea „conştienţei poate să-l reducă până şi pe cel mai pretenţios gânditor la o incoerenţă bâlbâită”. Motivul este că oamenii au adeseori concepţii neverificate şi inconsecvente cu privire la semnificaţiile acestui termen.

Mulţi observatori consideră conştienţa ca fiind o formă de performanţă – de exemplu, capacitatea de reflectare asupra sinelui, respectiv abilitatea de a-ţi

înţelege propriile gânduri şi de a le explica. Eu aş descrie conştienţa ca fiind abilitatea de a gândi asupra propriei gândiri. După câte s-ar putea presupune, am putea găsi o cale oarecare de evaluare a acestei abilităţi şi apoi de a folosi acest test pentru separarea lucrurilor conştiente de lucrurile inconştiente.

Pe de altă parte, însă, intrăm repede la încurcătură dacă încercăm să implementăm această abordare. Sugarul este conştient? Dar câinele? Niciunul şi nici celălalt nu sunt prea pricepuţi la descrierea propriilor procese de gândire. Există oameni care cred că sugarii şi câinii nu sunt fiinţe conştiente tocmai din cauză că nu se pot autoexplica. Dar cum rămâne atunci cu computerul pe nume Watson? E posibil să fie pus şi el într-un mod de funcţionare în care să ne poată practic lămuri cum de a dat un anumit răspuns? Deoarece conţine un model al propriei gândiri, Watson este, prin urmare, conştient, în vreme ce sugarul şi câinele nu sunt?

Înainte de a continua analizarea acestei chestiuni, este important să reflectăm asupra celei mai importante distincţii legate de ea: Ce anume putem lămuri de pe urma ştiinţei versus ce rămâne cu adevărat o problemă de filosofie? Un punct de vedere este acela că filosofia reprezintă un fel de refugiu pentru întrebări care nu au fost încă supuse metodei ştiinţifice. Potrivit acestei perspective, din momentul în care o ştiinţă progresează suficient de mult pentru a rezolva un anumit set de chestiuni, filosofii pot trece la alte preocupări, până soseşte momentul în care ştiinţa le rezolvă şi pe ele. Această perspectivă este endemică în ceea ce priveşte chestiunea conştienţei şi mai ales întrebarea „Ce şi cine este conştient?”

Să ne gândim la aceste afirmaţii ale filosofului John Searle: „Ştim că creierul produce conştienţa prin mecanisme biologice specifice… Esenţial este să se recunoască faptul că conştienţa e un proces biologic, asemenea digestiei, lactaţiei, fotosintezei sau mitozei…

Creierul este o maşină – o maşină biologică, pentru a fi mai exact, dar, totuşi, o maşină. Prin urmare, primul pas este să se stabilească felul cum face creierul acest lucru şi apoi să se construiască o maşină artificială care să aibă acelaşi mecanism, la fel de eficient, pentru producerea conştienţei".1 Oamenii sunt adeseori surprinşi să vadă citate precum acesta, deoarece ei presupun că Searle e devotat protejării misterului conştienţei de către reducţionişti de felul lui Ray Kurzweil.

Filosoful australian David Chalmers (născut în anul 1966) a inventat termenul „problema grea a conştientei" pentru a descrie dificultatea identificării cu precizie a acestui concept în esenţă indescriptibil. Uneori, o expresie scurtă rezumă o întreagă şcoală de gândire atât de bine, încât devine emblematică (de exemplu, „banalitatea răului", a lui Hannah Arendt). Celebra formulare a lui Chalmers face şi ea foarte bine acest lucru.

Discutând despre conştientă, este foarte uşor să aluneci către cercetarea atributelor observabile şi măsurabile pe care le asociem cu faptul de a fi conştient, dar acestei abordări îi lipseşte însăşi esenţa ideii. Tocmai am menţionat conceptul de metacogniţie – ideea de a gândi asupra propriei gândiri – ca fiind o asemenea corelare a conştienţei. Alţi observatori combină inteligenţa emoţională sau inteligenţa morală cu conştienţa. Dar, din nou, abilitatea noastră de a exprima un sentiment de iubire, de a pricepe o glumă sau de a fi sexy conţine, pur şi simplu, tipuri de performanţe – impresionante şi inteligente, poate, dar, cu toate acestea, abilităţi ce pot fi observate şi măsurate (chiar dacă mai discutăm asupra felului cum să le evaluăm). Descoperirea modului în care creierul realizează acest gen de sarcini şi a fenomenelor ce se petrec în creier atunci când sunt îndeplinite asemenea sarcini constituie întrebarea „simplă" a lui Chalmers. Desigur, problema „simplă" este, de fapt, oricum altcumva şi reprezintă, poate, cea mai dificilă şi mai importantă ţintă ştiinţifică a epocii noastre. Problema „grea" a lui Chalmers

este atât de grea, încât e, mai presus de toate, inexprimabilă.

În sprijinul acestei distincţii, Chalmers introduce un experiment mintal care implică ceea ce el numeşte *zombies*. Zombie este o entitate care se poartă exact ca un om, dar pur şi simplu nu are nicio experienţă subiectivă – cu alte cuvinte, zombie nu are conştienţă. Chalmers susţine că, de vreme ce putem să îi concepem, zombii sunt cel puţin logic posibili. Dacă te-ai afla la un cocteil la care ar fi prezenţi atât oameni „normali”, cât şi *zombies*, cum ţi-ai da seama de deosebirile dintre ei? Poate că situaţia se aseamănă cu cocteilul la care chiar ai fost.

Mulţi oameni răspund la această întrebare afirmând că i-ar interoga pe indivizii pe care ar dori să-i evalueze în privinţa reacţiilor lor emoţionale la evenimente şi idei. Un zombie, cred ei, şi-ar trăda lipsa de experienţă subiectivă printr-o deficienţă manifestată în anumite tipuri de răspunsuri emoţionale. Dar un răspuns pe aceste coordonate pur şi simplu nu conţine ipotezele experimentului mintal. Dacă am întâlni o persoană neemoţională (de pildă un individ cu deficienţe emoţionale, aşa cum se întâmplă în anumite cazuri de autism), un avatar sau un robot care nu este convingător ca fiinţă umană emoţională, entitatea nu ar fi un zombie. Reţine, cititorule: potrivit ipotezei lui Chalmers, un zombie este absolut normal sub aspectul capacităţii sale de a răspunde, inclusiv sub cel al abilităţii de a reacţiona emoţional; doar că îi lipseşte experienţa subiectivă. Ideea de bază este că nu există vreun mijloc de identificare a unui zombie, deoarece, prin definiţie, în comportament nu există niciun indiciu aparent al naturii de zombie. Deci e vorba despre o distincţie fără diferenţe?

Chalmers nu încearcă să răspundă la întrebarea grea, dar oferă nişte probabilităţi. Una constă într-o formă de dualism în care, în lumea fizică, nu există conştienţă ca atare, ci mai curând ca realitate ontologică separată. Potrivit acestei formulări, ceea ce face un om se bazează

pe procesele din creierul lui. Deoarece creierul este închis din punct de vedere cauzal, putem explica pe deplin acţiunile unei persoane, inclusiv gândurile sale, prin intermediul proceselor din creierul său. Pe urmă, conştienţă există într-un alt domeniu sau cel puţin este o proprietate diferită de a proceselor din lumea fizică. Această explicaţie nu permite minţii (cu alte cuvinte, proprietăţii conştiente asociate creierului) să afecteze creierul în mod cauzal.

O altă probabilitate pe care o enunţă Chalmers şi care nu este distinctă sub aspect logic de concepţia lui despre dualism, fiind adeseori numită panprotopsihism, susţine că toate sistemele fizice sunt conştiente, deşi omul este mai conştient decât, să spunem, un întrerupător. Cu siguranţă, aş fi de acord că creierul omului are mult mai multe lucruri de care să fie conştient decât un întrerupător de lumină.

Concepţia mea, care e probabil o subşcoală a panprotopsihismului, sună aşa: conştienţă e o proprietate emergentă a unui sistem fizic complex. Din această perspectivă, câinele este şi el conştient, dar pe undeva mai puţin decât omul. Şi furnica are un anumit nivel de conştienţă, însă mai redus decât al câinelui. Colonia de furnici, pe de altă parte, se poate considera că are un nivel de conştienţă mai înalt decât furnica individuală; adică, în mod cert, colonia e mai conştientă decât o singură furnică. Potrivit acestei socoteli, un computer care imită cu succes complexitatea creierului uman ar avea aceeaşi conştienţă emergentă ca un om.

Un alt mod de a conceptualiza ideea de conştienţă o imaginează pe aceasta ca sistem care are „qualia”. Dar ce înseamnă quaiia? O definiţie a termenului este „experienţe conştiente”. Însă definiţia asta nu ne duce prea departe. Gândeşte-te la următorul experiment mintal: un specialist în neuroştiinţă este complet orb la culori – nu genul de orbire din cauza căreia cineva amestecă anumite culori, să spunem verde şi roşu (ca mine), ci, dimpotrivă, o afecţiune

din cauza căreia individul afectat trăieşte într-o lume exclusiv alb-negru. (într-o versiune extremă a acestui scenariu, persoana respectivă a crescut într-o lume în alb-negru şi nu a văzut niciodată vreo culoare. Ideea este că, în lumea persoanei respective, nu există culori.) însă persoana cu pricina a studiat amănunţit fizica cromatică – este conştientă că lungimea de undă a luminii roşii e de 700 nanometri – şi procesele neurologice ale unei persoane care trăieşte în mod normal experienţa culorilor; prin urmare, ştie mult mai multe despre felul în care creierul procesează culorile. Ştie despre culori mai multe decât cei mai mulţi oameni. Dacă ai vrea să ajuţi respectiva persoană încercând să-i explici cum este de fapt această trăire a experienţei „roşii", cum ai face-o? Poate că ar trebui să-i citeşti o parte din poezia *Roşu a* poetului nigerian Oluseyi Oluseun:

Roşu-i culoarea sângelui simbol al vieţii Roşu-i culoarea pericolului simbol al morţii

Roşu-i culoarea trandafirilor simbol al *frumuseţii Roşu-i culoarea iubiţilor simbol al unităţii*

Roşu-i culoarea roşiei simbol al sănătăţii Roşu-i culoarea focului simbol al dorinţei arzânde

Asta i-ar crea o idee destul de bună despre unele dintre asocierile pe care le fac oamenii cu culoarea roşie şi ar putea chiar să-i faciliteze persoanei susţinerea propriilor asocieri într-o conversaţie despre culori. („Da, îmi place la nebunie culoarea roşie, e atât de arzătoare şi de aprinsă, atât de periculos de frumoasă…") Dacă ar vrea, probabil că ar putea să convingă pe cineva că a trăit experienţa culorii roşii, însă nici toată poezia de pe lume nu i-ar permite să trăiască într-adevăr această experienţă.

În mod similar, cum i-ai explica senzaţia de scufundare în apă cuiva care nu a atins apa niciodată? Am fi iarăşi forţaţi să recurgem la poezie, dar realmente nu există nicio modalitate de comunicare a experienţei propriu-zise. Aceste experienţe constituie ceea ce numim *qualia*.

Mulţi dintre cititorii acestei cărţi au trăit experienţa culorii roşii. Oarde unde pot eu să ştiu dacă experienţa culorii roşii pe care ai trăit-o tu nu este aceeaşi cu experienţa pe care o trăiesc eu când mă uit la culoarea albastră? Ne uităm amândoi la un obiect roşu şi afirmăm cu certitudine că este roşu, dar asta nu răspunde la întrebarea noastră. E posibil ca eu să am aceeaşi experienţă pe care o ai tu când te uiţi la culoarea albastră, dar am învăţat cu toţii să spunem că roşul e roşu. Am putea începe să apelăm iar la poezii, dar ele nu fac altceva decât să reflecte asocierile pe care le realizează oamenii pentru diverse culori; nu vorbesc despre natura de fapt a qualiei. Pe bună dreptate, oamenii orbi congenital au citit o mulţime de lucruri despre culori, asemenea referiri fiind extrem de abundente în literatură şi, prin urmare, acei oameni au o anumită versiune a unei experienţe a culorii. Cum se compară experienţa trăită de ei pentru culoarea roşu cu experienţa trăită de văzători? Pe bună dreptate, este aceeaşi întrebare ca în cazul femeii care trăieşte într-o lume alb-negru. E remarcabil faptul că fenomene atât de comune din viaţa noastră sunt complet inexprimabile şi fac imposibilă o confirmare simplă, cum ar fi aceea că trăim experienţa aceleiaşi *qualia*.

O altă definiţie pentru *qualia o* echivalează pe aceasta cu senzaţia dată de o experienţă. Însă definiţia nu e cu nimic mai puţin circulară decât încercările noastre de mai devreme de definire a conştienţei: expresiile „îmi dă senzaţia”, „Trăiesc o experienţă” şi „Conştienţă” sunt toate sinonime. Conştienţa şi chestiunea strâns legată de *qualia* reprezintă o problemă fundamentală, poate întrebarea filosofică supremă (deşi subiectul identităţii poate să fie chiar mai important, după cum voi arăta în secţiunea de încheiere a acestui capitol).

Deci, în privinţa conştienţei, care este întrebarea? Aceasta: Cine sau ce este conştient? în titlul acestei cărţi, mă refer la „minte” şi nu la „creier”, deoarece mintea este un creier conştient. Am putea spune de asemenea că

mintea are liber arbitru şi identitate. Afirmaţia că aceste subiecte sunt filosofice nu este axiomatică. Susţin că ele nu pot fi niciodată pe deplin decise prin intermediul ştiinţei. Cu alte cuvinte, nu există experimente falsificabile pe care să le putem avea în vedere pentru a rezolva problema fără a formula ipoteze filosofice. Dacă am construi un detector de conştienţă, Searle ar vrea ca dispozitivul să-l asigure că e plin de neurotransmiţători biologici. Filosoful american Daniel Dennett (născut în 1942) ar fi mai flexibil în privinţa substratului, dar ar putea dori să stabilească dacă sistemul conţine sau nu un model al lui însuşi şi al propriei performanţe. Această concepţie se apropie mai mult de a mea, dar în esenţa ei se află în continuare o ipoteză filosofică.

Au fost avansate în permanenţă propuneri de teorii ştiinţifice care leagă conştienţă de câte un atribut fizic măsurabil. La acesta Searle se referă denumindu-l „mecanism pentru crearea conştienţei”. Omul de ştiinţă, filosoful şi anesteziologul american Stuart Hameroff (născut în anul 1947) scria că „filamentele citoscheletale sunt rădăcinile conştienţei”.2 El se referă la firişoarele fine din fiecare celulă (inclusiv neuroni, dar nu numai ei) denumite minitubuli care conferă fiecărei celule integritate structurală şi joacă un rol în diviziunea celulară. Cărţile şi lucrările lui pe această temă conţin descrieri detaliate şi ecuaţii ce explică plauzibilitatea ca microtubulii să joace un rol în procesarea informaţiei la nivelul celulelor. Dar legătura microtubulilor cu conştienţă necesită un salt al credinţei care nu este fundamental diferit de saltul credinţei în cazul unei doctrine religioase ce descrie o fiinţă supremă care depozitează conştienţă (uneori denumită şi „suflet”) în anumite entităţi (de obicei umane). Se aduc unele dovezi slabe în sprijinul concepţiei lui Hameroff, în speţă fiind vorba despre observaţia că procesele neurobiologice care pot să susţină această presupusă computaţie celulară sunt stopate în timpul anesteziei. Dar asta e departe de a fi o demonstraţie

convingătoare, dat fiind faptul că o mulţime de procese sunt oprite în timpul anesteziei. Nu putem spune cu certitudine nici măcar că subiecţii nu sunt conştienţi sub anestezie. Tot ceea ce ştim e că oamenii nu-şi amintesc, pe urmă, experienţele. Nici măcar acest lucru nu este universal valabil, deoarece unele persoane pot rememora – cu precizie – experienţele din timpul anesteziei, inclusiv, de exemplu, conversaţiile chirurgilor. Denumit conştientizare sub anestezie, acest fenomen apare în Statele Unite de aproximativ de 40.000 de ori pe an.3 însă chiar şi lăsând asta la o parte, conştienţă şi memoria sunt concepte complet diferite. După cum am arătat pe larg, dacă mă gândesc la experienţele mele de peste zi minut-cu-minut, constat că am avut o mulţime de impresii senzoriale, dar îmi amintesc de doar foarte puţine dintre ele. Nu am fost conştient de ce am văzut sau auzit toată ziua? E o întrebare bună, dar răspunsul nu este prea clar.

Fizicianul şi matematicianul englez Roger Penrose (născut în anul 1931) a făcut un alt salt de credinţă în propunerea sursei conştientei, cu toate că şi teoria lui avea legătură cu microtubulii – mai exact, cu capacităţile lor de computare cuantică. Raţionamentul lui Penrose, deşi nu a fost afirmat explicit, a părut a fi acela că, la fel ca un efect cuantic, conştienţa este misterioasă, astfel încât cele două trebuie să fie legate într-un fel oarecare.

Penrose şi-a început analiza cu teoremele lui Turing referitoare la problemele de nerezolvat şi cu Teorema incompletitudinii a lui Godel. Premisa lui Turing (care a fost discutată foarte detaliat în Capitolul 8) spune că există probleme algoritmice care pot fi formulate, dar care nu pot fi rezolvate de o maşină Turing. Dată fiind universalitatea computaţională a maşinii Turing, putem concluziona că aceste „probleme insolvabile” nu pot fi rezolvate de nicio maşină. Teorema incompletitudinii a lui Godel ajunge la un rezultat foarte asemănător în privinţa abilităţii de a demonstra ipoteze ce implică numere. Argumentul lui Penrose susţine că creierul uman este capabil să rezolve

aceste probleme de nerezolvat, deci e capabil şi să facă lucruri pe care o maşină deterministă, cum ar fi un computer, nu le poate face. Ideea ce derivă din argumentul lui Penrose este, măcar parţial, aceea că fiinţa umană se află deasupra maşinilor. Dar premisa centrală – că omul poate rezolva problemele de nerezolvat ale lu' Turing şi Godel – este pur şi simplu neadevărată, din păcate.

O celebră problemă de nerezolvat, numită problema *busybeaver*, este formulată după cum urmează: Găseşte numărul maxim de 1 pe care îl poate scrie pe banda ei o maşină Turing cu un anumit număr de stări. Deci, pentru determinarea valorii *busybeaver* pentru numărul n, construim toate maşinile Turing care au n stări (un număr finit dacă n este finit) şi apoi determinăm cel mai mare număr de 1 pe care aceste maşini l-ar putea scrie pe benzile lor, excluzând maşinile Turing care intră într-o buclă infinită. Problema este de nerezolvat deoarece, atunci când încercăm să simulăm toate aceste maşini Turing cu n stări, simulatorul nostru va intra într-o buclă infinită în clipa în care va începe să simuleze una dintre maşinile Turing care intră într-o buclă infinită. Pe de altă parte însă, se dovedeşte că există computere care au reuşit totuşi să determine funcţia *busybeaver* pentru anumite valori ale lui *n*. La fel şi oamenii, dar computerele au rezolvat problema pentru mult mai multe n-uri decât oamenii neasistaţi. Computerele sunt în general mai bune decât omul la rezolvarea problemelor de nerezolvat ale lui Turing şi Gódel.

Penrose a legat aceste presupuse capacităţi transcendentale ale creierului uman de computarea cuantică din interiorul lui. Potrivit lui Penrose, aceste efecte cuantice neuronale sunt cumva inerent inaccesibile computerelor, prin urmare gândirea umană are un avantaj inerent. De fapt, electronica obişnuită foloseşte efecte cuantice (tranzistorii se bazează pe *turmel*ingul cuantic al electronilor prin bariere); computarea cuantică la nivelul creierului nu a fost demonstrată niciodată; performanţa

mintală a omului poate fi explicată satisfăcător cu ajutorul metodelor clasice de computare; şi, în orice caz, nimic nu ne împiedică să aplicăm computarea cuantică la computere. Nici măcar una dintre aceste obiecţii nu i-a fost adresată vreodată lui Penrose. Abia când criticii au subliniat că creierul este un loc prea cald şi dezordonat pentru computarea cuantică, Hameroff şi Penrose şi-au unit forţele. Penrose a găsit în neuroni un vehicul despre care se poate imagina că ar putea să susţină computarea cuantică – şi anume microtubulii, legat de care Hameroff speculase că fac parte din procesarea informaţiei la nivelul neuronului. Aşadar, teza Hameroff-Penrose afirmă că microtubulii din neuroni realizează computarea cuantică şi că această computare este răspunzătoare de conştienţă.

Teza celor doi a fost şi ea criticată, de exemplu de fizicianul şi cosmologul american de origine suedeză Max Tegmark (născut în anul 1967), care a determinat faptul că evenimentele cuantice din microtubuli pot supravieţui doar 10 – 13 microsecunde, ceea ce înseamnă o perioadă de timp prea scurtă şi pentru computarea unor rezultate cât de cât semnificative, şi pentru afectarea proceselor neuronale. Există anumite tipuri de probleme în cazul cărora computarea cuantică ar dovedi capacităţi superioare în raport cu computarea clasică – de exemplu spargerea codurilor prin factorizarea unor numere mari. Pe de altă parte, gândirea umană neasistată se dovedeşte a fi îngrozitoare în rezolvarea acestor probleme şi nu poate egala nici măcar computerele clasice în acest domeniu, fapt. ce sugerează că creierul nu demonstrează niciun fel de capacităţi de computare cuantică. Mai mult de-atât, chiar dacă un fenomen cum este computarea cuantică la nivelul creierului ar exista cu adevărat, el nu ar fi legat de conştienţă în mod obligatoriu.

Cum se construieşte o minte 199
Trebuie să ai credinţă i

Ce lucru desăvârşit e omul! Ce aleasă îi e cugetarea!

Şi cât de nemăsurate însuşirile! Ce falnic şi minunat e în portul şi mişcările lui! întocmai ca un înger în fapte! Şi ca un zeu în năzuinţe! Splendoarea lumii! Pilda perfecţiunii vietăţilor! Şi, totuşi, ce e pentru mine întruchiparea asta din lut?

— Hamlet, în *Hamlet* de Shakespeare

Realitatea este că aceste teorii sunt toate salturi ale credinţei. Aş adăuga că, în ceea ce priveşte conştienţa, principiul călăuzitor spune că „trebuie să ai credinţă" – cu alte cuvinte, fiecare dintre noi are nevoie de un salt al credinţei în privinţa a cine şi ce este conştient şi a ce şi cine suntem noi ca fiinţe conştiente. Altminteri, nu ne-am putea trezi dimineaţa. Dar ar trebui să fim cinstiţi cu privire la necesitatea fundamentală a unui salt al credinţei în această chestiune şi autoreflexivi în privinţa a ceea ce implică acest salt deosebit al nostru.

Oamenii fac salturi foarte diferite, în ciuda impresiilor de sens contrar. Premisele filosofice individuale privitoare la natura şi sursa conştienţei stau la baza dezacordurilor pe subiecte care merg de la drepturile animalelor la avort şi se concretizează în viitoare conflicte chiar şi mai litigioase pe subiectul drepturilor maşinilor. Predicţia mea obiectivă este că maşinile din viitor vor părea a fi conştiente şi că oamenilor biologici li se vor părea convingătoare atunci când vor vorbi despre percepţiile lor în condiţiile experienţelor subiective, conştiente.[1] Maşinile vor prezenta întreaga gamă a indiciilor subtile, familiar emoţionale: ne vor face să râdem şi să plângem şi se vor înfuria pe noi dacă le vom spune că nu credem că sunt conştiente. (Vor fi foarte inteligente, aşa că nu vom dori să se întâmple aşa ceva.) Vom ajunge să acceptăm că sunt persoane conştiente. Saltul meu personal de credinţă este următorul: din momentul în care

1 [în original, qualia] Termen folosit în filosofie cu sensul de instanţe Individuale ale experienţei subiective, conştiente. Termenul provine din latinescul qualia, care înseamnă „ce fel" ori „ce gen". Exemple de qualia: suferinţa provocată de o durere de cap, gustul vinului sau senzaţia de roşeaţă a cerului de seară, (nota trad.)

maşinile vor reuşi să fie convingătoare atunci când se vor referi la percepţiile lor în condiţiile experienţelor subiective, conştiente, vor fi realmente persoane conştiente. Am ajuns la această poziţie prin următorul experiment mintal: imaginează-ţi că te întâlneşti cu o entitate în viitor (un robot sau un avatar) care e complet convingătoare în reacţiile sale emoţionale. Râde convingător la glumele tale şi la rându-i te face să râzi şi să plângi (însă nu doar ciupindu-te). Te convinge de sinceritatea sa când îţi vorbeşte despre temerile şi dorinţele sale. Din toate punctele de vedere, entitatea pare conştientă. De fapt, pare să fie o persoană. Vei accepta că este o persoană conştientă?

Dacă reacţia ta iniţială este aceea că mai curând ai detecta vreun fel în care să-şi trădeze natura nonbiologică, atunci nu respecţi ipotezele acestei situaţii ipotetice, care au stabilit că este pe deplin convingătoare. Dată fiind această ipoteză, dacă ar fi ameninţată de distrugere şi ar răspunde aşa cum ar face-o un om, arătându-se terorizată, ai reacţiona în acelaşi mod empatic pe care l-ai manifesta dacă ai fi spectatorul unei scene în care ar fi implicat un om? în cazul meu, răspunsul este da şi cred că răspunsul ar fi acelaşi pentru cei mai mulţi, dacă nu pentru toţi ceilalţi oameni, indiferent de ce ar susţine acum în cadrul unei dezbateri filosofice. Din nou, accentul, aici, cade pe cuvântul „convingătoare”.

În mod cert, există dezacorduri privitoare la când sau dacă vom întâlni vreodată o asemenea entitate nonbiologică. Predicţia mea consecventă este că acest lucru se va petrece pentru prima oară în anul 2029 şi va deveni un fapt obişnuit în anii ’2030. Dar, lăsând deoparte cadrul timpului, cred că vom ajunge în cele din urmă să considerăm că asemenea entităţi sunt conştiente. Să ne gândim cum le tratăm deja când avem de-a face cu ele în calitate de personaje ale poveştilor şi filmelor: R_2D_2 din seria *Războiul Stelelor*, David şi Teddy din filmul *A.I.*, Data din serialul de televiziune *Star Trek*: *The Next Generation*,

Johnny 5 din filmul *Circuit*, WALL-E din filmul Disney *Wall-E*, T-800 – Terminatorul (cel bun) – din al doilea şi ultimul film*âermin ator*, Rachael Replica din filmul *Blade Runner* (care, apropo, nu este conştientă că nu e om), Bumblebee din filmul TV şi seria comică *Transformers* şi Sonny din filmul *I, Robot*. Empatizăm cu aceste personaje deşi ştim că sunt nonbiologice. Le considerăm ca fiind persoane conştiente, la fel cum facem cu personajele umane biologice. Le împărtăşim sentimentele şi ne temem pentru ele când intră în buclucuri. Dacă astăzi tratăm aşa personajele nonbiologice de ficţiune, atunci tot aşa vom trata în viitor şi inteligenţele din viaţa reală care se va întâmpla să nu aibă un substrat biologic.

Dacă accepţi saltul de credinţă că o entitate nonbiologică pe care o găseşti convingătoare în reacţiile sale faţă de experienţele subiective, conştiente, este într-adevăr conştientă, gândeşte-te doar la ce presupune asta: respectiv, că această conştientă este o proprietate emergentă a modelului de ansamblu al entităţii, nu substratul pe baza căruia funcţionează.

Există o prăpastie conceptuală între ştiinţă, care susţine măsurarea obiectivă şi concluziile pe care le tragem în acest fel, şi conştiinţă, care este un sinonim pentru experienţa subiectivă. În mod evident, nu putem să-i adresăm unei entităţi puse la îndoială întrebarea „Eşti conştientă?” Dacă ne-am uita în „capul” ei – biologic sau ba – ca să ne asigurăm de răspuns, am avea de făcut presupuneri filosofice pentru a determina ce anume căutăm. Întrebarea dacă o entitate e sau nu conştientă nu este, prin urmare, una ştiinţifică. Pe această bază, unii observatori merg mai departe şi pun sub semnul îndoielii faptul că conştiinţa însăşi are vreo bază în realitate. Scriitorul şi filosoful englez Susan Blackmore (născută în anul 1951) vorbeşte despre „marea iluzie a conştienţei”. Ea recunoaşte realitatea memei (ideea) conştienţei – cu alte cuvinte, conştienţa există într-adevăr ca idee şi există foarte multe structuri neocorticale care se ocupă de idee,

ca să nu mai vorbim de cuvintele care au fost rostite şi scrise despre ea. Dar nu este limpede dacă ea se referă la ceva real. Blackmore merge mai departe şi explică faptul că ea nu neagă neapărat realitatea conştienţei, ci mai curând încearcă să articuleze genurile de dileme cu care ne confruntăm atunci când încercăm să precizăm conceptul. După cum a scris psihologul şi autorul britanic Stuart Sutherland (1927 – 1998) în *International Dictionary of Psychology – Dicţionarul internaţional de psihologie*, „conştienţa este un fenomen fascinant, dar evaziv; este imposibil de specificat ce anume este, ce face sau de ce a evoluat”.4

În orice caz, ar fi bine să nu respingem conceptul cu prea mare uşurinţă, considerându-l doar o dezbatere politicoasă între filosofi – dezbatere care, întâmplător, datează de două mii de ani, de la dialogurile platoniene. Ideea conştienţei se află în sistemul nostru moral, iar sistemul nostru juridic, la rândul lui, este construit în mare pe aceste credinţe morale. Dacă cineva stinge conştienţa cuiva, cum se întâmplă în cazul unei crime, considerăm că acest act este Imoral şi, cu unele excepţii, chiar o mare crimă. Aceste excepţii sunt relevante şi pentru conştienţă, sub aspectul că putem autoriza forţele poliţieneşti sau militare să omoare anumiţi oameni conştienţi pentru protejarea unui număr mai mare de alţi oameni conştienţi. Putem dezbate meritele excepţiilor, dar principiul de bază rămâne valabil.

Se consideră de asemenea că atacarea unui om şi obligarea lui să îndure experienţa suferinţei este ceva imoral şi ilegal. A-mi distruge proprietatea este probabil un gest acceptabil. Dacă îţi distrug ţie proprietatea fără permisiunea ta, probabil că gestul nu este acceptabil, pentru că provoc suferinţă nu proprietăţii tale, ci ţie, în calitatea ta de deţinător al proprietăţii. Pe de altă parte, dacă proprietatea mea include o fiinţă conştientă, ca de pildă un animal, atunci eu, în calitate de deţinător al acelui animal, nu am neapărat libertatea morală sau legală să fac

cu el orice vreau – există, de exemplu, legi împotriva cruzimii faţă de animale.

Deoarece mare parte din sistemul nostru moral şi juridic se bazează pe protejarea existenţei şi prevenirea suferinţelor inutile provocate entităţilor conştiente, pentru a face judecăţi responsabile trebuie să răspundem la întrebarea „Cine este conştient?” Prin urmare, această întrebare nu este doar o chestiune de dezbatere intelectuală, fapt evident în cazul controversei care învăluie un subiect cum ar fi avortul. Ar trebui să subliniez că subiectul avortului poate să depăşească pe undeva tema conştienţei, partizanii proviaţă susţinând că potenţialul ca embrionul să devină în cele din urmă o persoană conştientă este un motiv suficient pentru a i se acorda acestuia protecţie, la fel cum şi o persoană aflată în comă are acest drept. Dar, la nivelul fundamental, subiectul este o dezbatere cu privire la momentul în care fătul devine conştient.

Percepţia conştienţei ne afectează adeseori şi judecata în domenii controversate. Revenind la subiectul avortului, mulţi oameni fac distincţie între o măsură cum ar fi pilula din dimineaţa de după, care împiedică implantarea embrionului în uter în primele zile ale sarcinii, şi avortul în stadiul ulterior. Distincţia se referă la posibilitatea ca fătul în stadiul ulterior să aibă conştienţă. Este dificil să susţii că un embrion cu vârsta de câteva zile este conştient dacă nu adopţi o poziţie panprotopsihistă, dar, chiar şi aşa, embrionul s-ar plasa sub cel mai simplu animal în termeni de conştiinţă. În mod similar, avem reacţii foarte diferite la maltratarea maimuţelor mari, comparativ cu uciderea insectelor, de exemplu. Nimeni nu-şi face prea multe griji astăzi că ar provoca dureri şi suferinţe software-ului computerului (deşi comentăm amănunţit capacitatea software-ului de a ne provoca nouă suferinţe), dar, când software-ul viitorului va avea inteligenţa intelectuală, emoţională şi morală a omului biologic, această preocupare va deveni reală.

Prin urmare, poziţia mea este că voi accepta entităţile nonbiologice care sunt pe deplin convingătoare în reacţiile lor emoţionale drept persoane conştiente, iar predicţia mea este că societatea le va accepta de asemenea. Observă, cititorule, că această definiţie merge dincolo de entităţile care pot trece Testul Turing, ce impune stăpânirea limbajului uman. Acestea din urmă sunt suficient de asemănătoare omului ca să le pot include şi cred că şi cea mai mare parte a societăţii le va include, însă eu adaug aici şi entităţile care manifestă reacţii emoţionale umane, dar pot să nu fie în măsură să treacă Testul Turing – de exemplu, copiii mici.

Răspunde asta la întrebarea filosofică cine este conştient, cel puţin pentru mine şi pentru alţii care acceptă acest salt al credinţei? Răspunsul este: nu chiar. Am acoperit doar un caz, cel al entităţilor care acţionează de o manieră umană. Cu toate că discutăm despre viitoare entităţi, discutăm şi despre entităţi care demonstrează reacţii umane convingătoare, deci această poziţie este în continuare centrată pe om. Dar cum rămâne cu formele de inteligenţă mai deosebite care nu sunt asemănătoare omului? Ne putem imagina inteligenţe la fel de complexe sau poate mult mai complexe şi mai complicate decât creierul omului, dar care au emoţii şi motivaţii complet diferite. Cum să decidem dacă sunt sau nu conştiente?

Putem începe analizând fiinţe din lumea biologică al căror creier este comparabil cu ai omului, dar, cu toate acestea, manifestă genuri de comportamente cu totul diferite. Filosoful britanic David Cockburn (născut în anul 1949) scrie despre urmărirea unei filmări cu un calmar gigant atacat (sau care credea că este atacat – Cockburn a emis ipoteza că animalul s-ar fi putut teme de omul cu camera de filmat). Calmarul s-a cutremurat şi s-a ferit, iar Cockburn scrie: „A reacţionat într-un mod care m-a impresionat imediat şi puternic, deoarece era o manifestare a fricii. Partea şocantă în această succesiune a fost felul în care a fost posibil să văd în comportamentul

unei fiinţe atât de diferite fizic de om o emoţie care era, fără niciun fel de ambiguitate şi foarte precis, una de frică”.5 Autorul concluzionează că animalul a trăit această emoţie şi îşi exprimă credinţa că oamenii care au văzut acel film au ajuns la aceeaşi concluzie în majoritatea lor. Dacă acceptăm descrierea şi concluzia lui Cockburn, atunci ar trebui să adăugăm pe lista noastră de entităţi conştiente calmarul gigant. Pe de altă parte însă, nici aşa nu am ajuns prea departe, deoarece concluzia se bazează pe reacţia noastră empatică la o emoţie pe care o recunoaştem în noi înşine. Este o perspectivă egocentrică sau umanocentrică.

Dacă ieşim din domeniul biologiei, inteligenţa nonbiologică va fi chiar şi mai variată decât inteligenţa din lumea biologică. De exemplu, e posibil ca unele entităţi să nu se teamă de propria distrugere şi să nu resimtă nevoia de emoţii pe care o vedem la om sau la orice fiinţă biologică. Poate că asemenea entităţi vor reuşi să treacă Testul Turing sau poate că nici nu vor fi dornice să încerce.

De fapt, astăzi construim roboţi lipsiţi de simţul autoconservării pentru a realiza misiuni în medii periculoase. Încă nu sunt suficient de inteligenţi sau de complecşi pentru ca noi să fim obligaţi să ţinem seama de sensibilitatea lor, dar ne putem imagina viitori roboţi de acest gen la fel de complecşi ca omul. Ce vom face în cazul lor?

Personal, aş spune că, dacă aş vedea în comportamentul unui asemenea dispozitiv devotamentul faţă de un scop complex şi meritoriu şi abilitatea de a lua decizii şi a executa acţiuni notabile pentru îndeplinirea misiunii, aş fi impresionat şi probabil că m-aş simţi deranjat dacă el ar fi distrus. Poate că forţez puţin conceptul, în sensul că răspund la un comportament care nu include multe emoţii considerate universale la om şi chiar la fiinţele biologice de toate felurile. Însă, din nou, caut să mă conectez la atribute cu care pot relaţiona în mine şi în alţi oameni. Ideea unei entităţi total devotate

unui scop nobil şi îndeplinirii lui sau cel puţin încercării de a face acest lucru fără să se gândească la binele său nu-i este, la urma urmelor, deloc străină experienţei umane. În această clipă analizăm, de asemenea, o entitate care caută să protejeze oamenii biologici sau să ne promoveze cumva interesele.

Şi dacă această entitate îşi are propriile scopuri, distincte de ale omului, şi nu îndeplineşte o misiune pe care am recunoaşte-o ca fiind nobilă în termenii ce ne sunt proprii? Aş putea atunci să încerc să văd dacă aş putea lega de ceva sau aprecia unele dintre abilităţile ei într-un alt mod. Dacă într-adevăr este deosebit de interigentăi e posibil să fie bună la matematică, aşa că probabil aş putea să port cu ea o conversaţie pe acest subiect. Poate că mi-ar aprecia glumele matematice.

Dar dacă entitatea nu manifestă niciun interes să comunice cu mine, iar eu nu am suficient acces la acţiunile sale şi la luarea deciziilor pentru a fi impresionat de frumuseţea proceselor sale interne, conduc oare aceste lucruri la concluzia că entitatea nu este conştientă? Trebuie să concluzionez că entităţile care nu reuşesc să mă convingă de reacţiile lor emoţionale sau care nu sunt preocupate nici măcar să încerce s-o facă nu sunt în mod necesar lipsite de conştienţă. Ar fi dificil să recunosc o altă entitate conştientă fără să stabilesc un anumit nivel de comunicare empatică, dar această judecată reflectă propriile mele limitări mai mult decât limitările entităţii în discuţie. Prin urmare, trebuie să procedăm cu umilinţă. Este suficient de provocator să ne punem, subiectiv, în locul altui om, aşa încât sarcina e cu atât mai dificilă în cazul inteligenţelor care sunt extrem de diferite de a noastră.

De ce *anume suntem conştienţi*?

Dacă am putea privi prin craniu în creierul unei persoane care gândeşte în mod conştient şi dacă locul excitaţiei optime ar fi luminos, atunci ar trebui să vedem jucând pe suprafaţa cerebrală o pată strălucitoare cu

marginile fantastice, unduitoare, fluctuând permanent ca mărime şi formă, înconjurată de un întuneric mai mult sau mai puţin profund ce acoperă restul emisferei.

— Ivan Petrovici Pavlov, 1913 e

Revenind la calmarul gigant, putem recunoaşte unele dintre emoţiile sale aparente, dar cea mai mare parte din comportamentul său este un mister. Cum e să fii un calmar gigant? Cum se simte acesta când îşi strecoară corpul lipsit de coloană vertebrală printr-o deschizătură mică? Nici măcar nu avem vocabularul necesar pentru a răspunde la această întrebare, dat fiind faptul că nu putem descrie nici măcar experienţele pe care le împărtăşim cu alţi oameni, de exemplu culoarea roşie sau senzaţia apei care ne stropeşte corpurile.

Dar nu trebuie să ajungem tocmai pe fundul oceanului pentru a găsi mistere în natura experienţelor conştiente – este suficient să ne gândim la ale noastre. Eu ştiu, de pildă, că sunt conştient. Eu presupun că tu, cititorule, eşti de asemenea conştient. (în ce-i priveşte pe oamenii care nu mi-au cumpărat cartea, nu sunt chiar atât de sigur.) Dar de ce anume sunt eu conştient? Ai putea să-ţi pui şi tu, cititorule, această întrebare. Încearcă următorul experiment mintal (care va da rezultate în cazul acelora dintre voi care conduceţi automobilul): imaginează-ţi că te deplasezi pe banda stângă a autostrăzii. Acum închide ochii, pune mâinile pe volanul imaginar şi fă mişcările necesare pentru schimbarea benzii la dreapta, în regulă, înainte să continui să citeşti, încearcă, latăce-ai făcut, probabil: ai apucat volanul. Ai verificat dacă banda din dreapta este liberă. Presupunând că banda era liberă, ai rotit volanul la dreapta pentru scurt timp. Pe urmă, l-ai îndreptat la loc. S-a făcut.

E bine că nu ai fost într-o maşină adevărată, pentru că tocmai ai trecut peste toate benzile autostrăzii şi te-ai izbit de un copac. Poate că eu ar fi trebuit să-ţi spun să nu încerci asta într-un automobil real aflat în mişcare (dar presupun că, totuşi, ai înţeles deja regula că nu se conduce

cu ochii închişi), deşi nu aceasta este aici problema-cheie. Dacă ai folosit procedura pe care tocmai am descris-o – şi aproape toată lumea o face în cazul acestui experiment mintal –, ai priceput totul greşit. Dacă mişti volanul la dreapta şi apoi îl aduci la poziţia iniţială, faci maşina să se îndrepte într-o direcţie diagonală pe direcţia ei de la început. Traversezi benzile din dreapta exact aşa cum ai intenţionat, dar continui să mergi la infinit spre dreapta până când ieşi de pe drum. Ceea ce trebuia să faci în timp ce maşina schimba banda spre dreapta era să roteşti volanul la stânga tot atât cât l-ai rotit la dreapta, iar apoi să-l aduci la poziţia iniţială. Asta ar fi făcut ca maşina să meargă drept înainte pe noua bandă.

Gândeşte-te că, dacă eşti şofer profesionist, ai făcut aceste mişcări de mii de ori. Nu eşti conştient când faci acest lucru? N-ai fost niciodată atent la ce faci atunci când schimbi benzile în mod real? Presupunând că nu citeşti această carte în spital, în convalescenţă după un accident produs la schimbarea benzilor, în mod cert stăpâneşti această abilitate. Dar, cu toate astea, nu eşti conştient de ce-ai făcut, chiar dacă ai făcut această mişcare de nenumărate ori.

Când îşi povestesc experienţele, oamenii le descriu ca succesiuni de situaţii şi decizii. Dar nu aşa trăim experienţa unei povestiri la început. Experienţa noastră iniţială reprezintă o succesiune de modele de nivel superior, dintre care o parte e posibil să fi generat senzaţii. În cel mai bun caz, ne amintim doar o mică parte din aceste modele. Chiar dacă suntem rezonabil de precişi în relatarea unei poveşti, ne folosim puterea de fabulaţie pentru a completa detaliile-lipsă şi a transforma succesiunea într-o naraţiune coerentă. Nu putem şti cu siguranţă care a fost experienţa noastră conştientă din amintirea ei; memoria este singura poartă de acces care ne poate aduce din nou faţă în faţă cu acea experienţă. Momentul actual este, să zicem, flotant şi se transformă rapid într-o amintire sau, cel mai adesea, nici atât. Chiar

dacă o experienţă e transformată într-o amintire, aceasta este stocată, aşa cum indică Teoria minţii bazate pe recunoaşterea modelelor, ca model de nivel superior compus din alte modele, într-o ierarhie colosală. După cum am subliniat de mai multe ori, aproape toate experienţele pe care le trăim (cum ar fi ocaziile în care am schimbat benzile) sunt imediat uitate. Aşadar, certitudinea în privinţa a ceea ce constituie experienţa noastră conştientă este practic ceva de neatins.

Estul este est şi vestul este vest înainte să existe creierul, în Univers nu existau nici culori şi nici sunete, nici mirosuri şi nici arome şi probabil că nu ar fi existat nici prea multe simţuri fără senzaţii sau emoţii.

— Roger W. Sperry7

René Descartes intră într-un restaurant şi se aşază să ia cina. Ospătarul vine la el şi îl întreabă dacă ar dori un apetitiv.

— Nu, mulţumesc, spune Descartes. Aş vrea doar să comand cina.

— Doriţi să ştiţi care sunt specialităţile zilei? întreabă ospătarul.

— Nu, răspunde Descartes, care devine nerăbdător.

— Doriţi să beţi ceva înainte de cină? întreabă ospătarul.

Descartes se simte insultat, deoarece e abstinent.

— Nu prea cred! spune el indignat şi, poofl, dispare.

— O glumă, aşa cum mi-o amintesc, spusă de

David Chalmers

Există două moduri de a privi chestiunile pe care le-am luat în considerare – perspectivele corespondente, occidentale şi răsăritene, asupra naturii conştienţei şi realităţii. În perspectiva apuseană, pornim de la o lume fizică în care au fost dezvoltate modele de informaţie. După câteva miliarde de ani de evoluţie, entităţile din acea lume au evoluat suficient pentru a deveni fiinţe conştiente. În perspectiva răsăriteană, conştienţa este realitatea fundamentală; lumea fizică prinde viaţă doar prin

intermediul gândirii fiinţelor conştiente. Cu alte cuvinte, lumea fizică reprezintă gândurile devenite manifeste ale fiinţelor conştiente. Acestea sunt, desigur, doar simplificări ale îilosofiilor complexe şi diverse, dar întruchipează principalele polarităţi ale filosofiilor conştienţei şi ale relaţiilor acesteia cu lumea fizică. Disensiunea Răsărit-Occident pe tema conştienţei şi-a mai găsit expresia şi în şcolile antagoniste de gândire din domeniul fizicii subatomice. În mecanica cuantică, particulele există într-o formă denumită câmp de probabilitate. Orice măsurătoare la care sunt supuse cu ajutorul unor dispozitive construite în acest scop produce ceea ce se numeşte colapsul funcţiei de undă, asta însemnând că particula capătă, brusc, o anumită localizare. O concepţie larg răspândită precizează faptul că o asemenea măsurătoare implică observarea de către un observator conştient, deoarece, altminteri, măsurătoarea ar fi un concept lipsit de sens. Prin urmare, particula capătă o anumită localizare (ca şi alte proprietăţi, cum ar fi viteza) numai atunci când este observată. În esenţă, particulele se gândesc că, dacă tot nu e nimeni care să se uite la ele, nici ele nu trebuie să decidă unde se află. Eu numesc asta şcoala budistă de mecanică cuantică, deoarece, potrivit acesteia, particulele nu există înainte de a fi observate de o persoană conştientă.

Mai există o interpretare a mecanicii cuantice care evită o asemenea terminologie antropomorfică. În această analiză, câmpul ce reprezintă particula nu este unul de probabilitate, ci mai curând o funcţie care are valori diferite în locuri diferite. Prin urmare, câmpul se află, la modul fundamental, acolo unde se află şi particula. Există constrângeri la care valorile câmpului sunt supuse în diferite locuri, deoarece întregul câmp ce reprezintă o particulă acoperă doar un volum limitat de informaţie. De aici vine şi cuvântul „cuantic”. Aşa-numitul colaps al funcţiei de undă, susţine această interpretare, nu este nici pe departe un colaps. Funcţia de undă nu dispare practic niciodată. Doar că instrumentul de măsură este făcut şi ei

din particule cu câmpuri, iar interacţiunea câmpului particulei care e măsurată cu câmpurile particulelor instrumentului de măsură se concretizează prin citirea fapfului că particula s-ar afla într-un anumit loc. Însă câmpul este prezent în continuare. Aceasta e interpretarea apuseană a mecanicii cuantice, deşi merită observat faptul că punctul de vedere mai popular în rândurile fizicienilor din lumea întreagă îl reprezintă ceea ce am numit eu *interpretarea răsăriteană*.

A existat un filosof ale cărui scrieri au reparat această fractură Răsărit-Apus. Gânditorul britanic de origine austriacă Ludwig Wittgenstein (1889 – 1951) a studiat filosofia limbii şi a cunoaşterii şi a cugetat îndelung la întrebarea „Ce putem şti cu adevărat?" A cumpănit acest subiect ca soldat în Primul Război Mondial şi a făcut notiţe pentru ceea ce avea să fie singura sa carte publicată în timpul vieţii, *Tractatus Logico-Philosophicus*. Lucrarea avea o structură neobişnuită şi abia prin eforturile fostului său mentor, matematicianul şi filosoful britanic Bertrand Russell, şi-a găsit un editor în anul 1921. A devenit Biblia unei şcoli majore de filosofie, cunoscută sub denumirea de pozitivism logic, care urmărea să definească limitele ştiinţei. Cartea şi mişcarea care a înconjurat-o l-au influenţat pe Turing şi apariţia teoriei computaţiei şi lingvisticii. *Tractatus Logico-Phiiosophicus* intuieşte că întreaga cunoaştere este ierarhică. De fapt, cartea însăşi este construită din afirmaţii înseriate şi numerotate. De exemplu, primele patru afirmaţii ale cărţii sunt:

1. Lumea este tot ceea ce e cazul.

1.1. Lumea este totalitatea faptelor, nu a lucrurilor.

1.11. Lumea este determinată de fapte şi de faptul că toate sunt fapte.

1.12. Căci totalitatea faptelor determină ce e cazul şi, de asemenea, ce nu e cazul.

O altă afirmaţie semnificativă din *Tractatus* – şi cea căreia Turing i-a devenit ecoul – este aceasta:

4.0031. Întreaga filosofie este o critică a limbii.

În esenţă, atât *Tractatus Logico-Phiiosophicus*, cât şi mişcarea pozitivismului logic afirmă că realitatea fizică există separat de perceperea realităţii fizice pe care o deţinem noi, dar că tot ceea ce putem şti despre realitate e ceea ce percepem cu ajutorul simţurilor – care pot fi îmbogăţite cu ajutorul instrumentelor noastre – şi cu ajutorul deducţiilor logice pe care le putem realiza în urma acestor impresii senzoriale. În esenţă, Wittgenstein încearcă să descrie metodele şi scopurile ştiinţei.

Afirmaţia finală din carte are numărul 7: „Ceea ce nu putem vorbi trebuie transmis mai departe în tăcere”. În consecinţă, acel Wittgenstein de la început considera discuţia despre conştienţă ca fiind circulară şi tautologică – prin urmare, o risipă de timp.

Însă celălalt Wittgenstein, de mai târziu, respinge complet această abordare şi acordă întreaga sa atenţie filosofică discuţiei despre chestiunile în legătură cu care argumentase anterior că trebuie să fie transmise mai departe în tăcere. Lucrările lui pe tema acestei gândiri revizuite au fost adunate şi publicate în 1953, la doi ani după moartea filosofului, într-o carte intitulată *Philosophical Investigations* – *Investigaţii filosofice*. Şi-a criticat ideile anterioare din *Tractatus*, considerându-le circulare şi goale de conţinut, şi a ajuns la concepţia că lucrurile asupra cărora recomandase să se păstreze tăcerea erau, de fapt, chiar cele asupra cărora merită să reflectăm. Aceste scrieri i-au influenţat profund pe existenţialişti, făcând din Wittgenstein uni Gul personaj din filosofia modernă care a devenit arhitectul de prim rang al unui număr de două şcoli contradictorii de gândire filosofică.

Dar despre ce anume considera Wittgenstein cel de mai târziu că merită să se cumpănească şi să se vorbească? Despre subiecte precum frumuseţea şi iubirea, pe care le recunoştea ca existând sub formă imperfectă de idei în minţile oamenilor. În acelaşi timp însă, filosoful scrie că asemenea concepte există cu adevărat într-o sferă

perfectă şi idealizată, similară „formelor” perfecte despre care a scris Platon în *Dialoguri*, o altă lucrare care a adus lumină asupra abordărilor naturii realităţii, aparent contradictorie.

Un gânditor a cărui poziţie cred că e greşit etichetată este filosoful şi matematicianul francez René Descartes. Celebra sa formulă „Gândesc, deci exist” este interpretată la modul general ca o preamărire a gândirii raţionale, în sensul că „Eu gândesc, adică pot să exercit gândirea raţională, prin urmare sunt valoros”. Pe cale de consecinţă, Descartes este considerat arhitectul perspectivei raţionale apusene.

Pe de altă parte însă, citind această afirmaţie în contextul altor scrieri, am o altă impresie. Descartes a fost tulburat de ceea ce se numeşte „problema minte-corp”. Mai exact, cum este posibil să apară o minte conştientă din materia fizică a creierului? Potrivit acestei perspective, s-ar părea că încerca să împingă scepticismul raţional la punctul de fractură, aşa că, din punctul meu de vedere, afirmaţia lui înseamnă, de fapt, „Gândesc, respectiv se petrece o experienţă subiectivă, prin urmare tot ce ştiu cu siguranţă e că există ceva – să-i spunem Eu”. Nu a putut fi sigur că lumea fizică există, pentru că tot ce avem la dispoziţie sunt impresiile simţurilor individuale determinate de ea, care pot fi greşite sau complet iluzorii. Pe de altă parte, însă, ştim că există experimentatorul.

Am fost crescut în sânul Bisericii Unitariene, în care am studiat toate religiile lumii. Studiam, să spunem, şase luni religia budistă, apoi treceam la serviciul religios budist, citeam cărţile religiei budiste şi purtam discuţii de grup cu liderii budişti. Pe urmă ne mutam la o altă religie, cum ar fi iudaismul. Tema generală era „numeroasele căi către adevăr”, laolaltă cu toleranţa şi transcendenţa. Această din urmă idee însemna că rezolvarea aparentelor contradicţii dintre tradiţii nu impune luarea deciziei că una este corectă, iar cealaltă greşită. Adevărul poate fi descoperit numai găsindu-se o explicaţie care elimină –

transcende – aparentele diferenţe, în special în cazul întrebărilor fundamentale legate de semnificaţie şi scop.

Aşa rezolv eu diferendele Apus-Răsărit privitoare la conştienţă şi la lumea fizică. Din punctul meu de vedere, ambele perspective trebuie să fie adevărate.

Pe de o parte, este o prostie să se nege lumea fizică. Chiar dacă trăim într-o simulare, cum specula filosoful suedez Nick Bostrom, realitatea este, totuşi, pentru noi un nivel conceptual real. Dacă acceptăm existenţa lumii fizice şi evoluţia care s-a petrecut în sânul ei, atunci putem vedea că din ea au evoluat entităţi conştiente.

Pe de altă parte, perspectiva răsăriteană – potrivit căreia conştienţa este fundamentală şi reprezintă unica realitate cu adevărat importantă – e de asemenea dificil de negat. Gândeşte-te numai, cititorule, la respectul deosebit pe care îl acordăm persoanelor conştiente faţă de cel acordat lucrurilor neconştiente. Considerăm că acestea din urmă nu au nicio valoare intrinsecă în afara măsurii în care pot influenţa experienţa subiectivă a persoanelor conştiente. Chiar dacă înţelegem conştienţa ca pe o proprietate emergentă a unui sistem complex, nu putem spune că ar fi vorba doar de un alt atribut (laolaltă cu „digestia” şi „lactaţia”, ca să-l citez pe John Searle). Reprezintă ceea ce e cu adevărat important.

Cuvântul „spiritual” este adeseori folosit pentru a denumi lucruri cu o semnificaţie supremă. Multor oameni nu le place să folosească o asemenea terminologie provenită din tradiţiile spirituale sau religioase pentru că implică un set de credinţe cu care s-ar putea să nu fie de acord. Dar, dacă eliminăm complexitatea mistică a tradiţiilor religioase şi respectăm pur şi simplu termenul „spiritual” ca pe unul care implică un fapt cu semnificaţie profundă pentru om, conceptul de conştienţă este foarte potrivit. El reflectă suprema valoare spirituală. Pe bună dreptate, însuşi „spiritul” este adeseori folosit pentru a denumi conştienţa.

Atunci, evoluţia poate fi considerată un proces

spiritual prin faptul că creează fiinţe spirituale, respectiv entităţi conştiente. Iar evoluţia se îndreaptă către o şi mai mare complexitate, o şi mai mare cunoaştere, o şi mai mare inteligenţă, o şi mai mare frumuseţe, o şi mai mare creativitate şi abilitate de a exprima emoţii mai transcendentale, cum ar fi iubirea. Toate acestea sunt descrieri pe care oamenii le-au folosit pentru conceptul de Dumnezeu, cu toate că Dumnezeu este descris ca neavând limite din aceste puncte de vedere.

Oamenii se simt adeseori ameninţaţi de discuţii care presupun posibilitattea ca o maşină să fie conştientă, deoarece au impresia că orice considerent în acest sens ar fi o denigrare a valorii spirituale a persoanei conştiente. Dar această reacţie reflectă o înţelegere greşită a conceptului de maşină. Criticii de acest gen abordează chestiunea pe baza maşinilor pe care le cunosc acum şi, oricât de impresionante ar deveni acestea, sunt de acord că exemplele tehnologice contemporane nu sunt încă demne de respectul nostru ca fiinţe conştiente. Predicţia mea este că vor deveni de nedistins de omul biologic, pe care îl considerăm într-adevăr o fiinţă conştientă şi, prin urmare, se vor împărtăşi din valoarea spirituală pe care o atribuim conştienţei. Nu este vorba despre o discreditare a omului; dimpotrivă, este o înălţare a felului în care înţelegem (unele dintre) maşinile viitoare. Probabil că ar trebui să adoptăm o terminologie diferită pentru aceste entităţi, de vreme ce vor fi maşini de un alt gen.

Pe bună dreptate, când privim acum în interiorul creierului şi îi decodificăm mecanismele, descoperim metode şi algoritmi pe care nu-i putem doar înţelege, ci îi putem şi recrea – „părţi ale unei prese care se împing unele pe altele”, ca să-l parafrazez pe matematicianul şi filosoful german Gottfried Wilheim Leibniz (1646 – 1716) când a scris despre creier. Omul alcătuieşte deja maşini spirituale. Mai mult de-atât, ne vom contopi cu instrumentele pe care le creăm într-atât, încât distincţia dintre om şi maşină se va estompa până când va dispărea.

Acest proces este deja în curs de desfăşurare, chiar dacă majoritatea maşinilor care ne devin prelungiri încă nu se află în interiorul trupurilor şi creierelor noastre.

Liberul arbitru

Un aspect central al conştienţei este capacitatea de a privi înainte, pe care o numim „anticipare”. Este capacitatea de a planifica şi, în termeni sociali, de a schiţa un scenariu pentru ceea ce este posibil să se întâmple sau se va întâmpla în mod cert în interacţiunile sociale care nu s-au produs încă. Prin acest sistem ne îmbunătăţim şansele de a face acele lucruri care vor reprezenta interesul nostru… Sugerez că „liberul arbitru” reprezintă aparenta noastră capacitate de a alege şi acţiona aşa cum ni se pare cel mai util sau apropiat şi, totodată, insistenţa noastră că aceste alegeri ne aparţin.

— Richard D. Alexander

Vom spune că planta nu ştie ce se petrece pentru simplul motiv că nu are ochi, urechi sau creier? Dacă spunem că acţionează mecanic sau numai mecanic, nu vom fi forţaţi să admitem că şi alte felurite acţiuni aparent deliberate sunt de asemenea mecanice? Dacă nouă ni se pare că planta ucide şi mănâncă o muscă în mod deliberat, plantei nu ar trebui să i se pară că omul trebuie să ucidă şi să mănânce o oaie în mod mecanic?

— Samuel Butler, 1871

Este creierul, care e remarcabil de dublu ca structură, un organ dublu, „aparent separat, dar totuşi o unitate în separaţie”?

— Henry Maudsley8

Redundanţa, după cum am aflat, este o strategie-cheie folosită de neocortex. Dar în creier mai există un nivel de redundanţă, în sensul că emisferele sale stângă şi dreaptă, deşi nu sunt identice, sunt în mare măsură similare. La fel cum anumite regiuni ale neocortexului ajung în mod normal să proceseze anumite tipuri de informaţii, emisferele se specializează şi ele – de exemplu, emisfera stângă e răspunzătoare de obicei de limbajul

verbal. Dar aceste alocări pot să fie şi rerutate, până în punctul în care putem supravieţui şi funcţiona oarecum normal cu o singură jumătate de creier. Steila de Bode şi Susan Curtiss, cercetători americani în domeniul neuropsihologiei, au comunicat un număr de patruzeci şi nouă de copii care au fost supuşi unei hemisferotomii (extirparea unei jumătăţi din creier), intervenţie chirurgicală extremă, aplicată pacienţilor care suferă de atacuri cerebrale ce le pun viaţa în pericol şi se manifestă ia nivelul unei singure emisfere. Unii copii supuşi acestei proceduri rămân cu deficienţe, dar deficienţele sunt specifice, iar pacienţii au personalităţi rezonabil de normale. Mulţi dintre ei se dezvoltă, aşa încât celor care îi văd nu le este deloc evident că au doar o jumătate de creier. De Bode şi Curtiss scriu despre copii hemisferotomizaţi pe stânga care „îşi dezvoltă un limbaj remarcabil de bun, în ciuda extirpării «emisferei limbajului»”.9 Autoarele descriu un asemenea student care a terminat colegiul, a urmat cursuri postuniversitare şi a obţinut scoruri peste medie la testele IQ. Studiile efectuate au indicat efecte minime pe termen lung asupra cogniţiei de ansamblu, memoriei, personalităţii şi simţului umorului.10 în cursul unui studiu realizat în anul 2007, cercetătorii americani Shearwood McClelland şi Robert Maxwell au adus la lumină rezultate la fel de pozitive pe termen lung la adulţi.11

S-a comunicat, de asemenea, că o fetiţă de zece ani din Germania, care s-a născut cu o jumătate de creier, era cât se poate de normală. Avea vederea aproape perfectă pentru un ochi, în vreme ce pacienţii care sunt supuşi unei interyenţii chirurgicale de hemisferotomie îşi pierd o parte din câmpul vizual imediat după operaţie.12 Cercetătorul scoţian Lars Muckli a comentat: „Creierul are o plasticitate uluitoare, dar am fost cât se poate de uluiţi când am văzut cât de bine s-a adaptat emisfera stângă a creierului acestei fetiţe pentru compensarea jumătăţii-lipsă”.

Deşi aceste observaţii sprijină cu siguranţă ideea

plasticităţii neocortexului, implicaţia lor mai interesantă este aceea că fiecare părem a avea două creiere, nu unul singur, şi că ne putem descurca foarte bine cu oricare dintre ele. Dacă îl pierdem pe unul, pierdem modelele corticale care sunt stocate exclusiv în el, dar fiecare creier este, în sine, cât se poate de complet. Prin urmare, fiecare emisferă îşi are o conştienţă proprie? Se poate argumenta că aşa stau lucrurile.

Să ne gândim la pacienţii cu scindare cerebrală, care au în continuare cele două emisfere ale creierului, dar canalul dintre ele este secţionat. Corpul calos e un mănunchi format din aproximativ 250 de milioane de axoni care interconectează emisferele cerebrale stângă şi dreaptă şi le permite acestora să comunice şi să se coordoneze una cu cealaltă. La fel cum doi oameni pot să comunice între ei strâns şi să acţioneze ca un singur factor de decizie, deşi rămân doi indivizi separaţi şi completi, cele două emisfere ale creierului pot funcţiona ca o unitate, rămânând totuşi independente.

Aşa cum sugerează şi termenul, la pacienţii care suferă de scindare cerebrală corpul calos a fost tăiat sau mutilat, aceştia rămânând efectiv cu două creiere funcţionale, fără comunicare directă între ele. Michael Gazzaniga, cercetător american din domeniul psihologiei (născut în anul 1939), a realizat experimente extensive pentru a afla ce gândeşte fiecare emisferă cerebrală a unui pacient cu scindare cerebrală.

Emisfera stângă a unui asemenea pacient vede în general câmpul vizual drept şi invers. Gazzaniga şi colegii săi i-au arătat unui subiect cu scindare cerebrală imaginea unei gheare de pui în câmpul vizual drept (imagine care era văzută de emisfera lui stângă) şi o scenă ceţoasă în câmpul vizual stâng (care era văzută de emisfera dreaptă). Pe urmă i-au prezentat o colecţie de fotografii, astfel încât acestea să poată fi văzute de ambele emisfere, şi i-au cerut pacientului să aleagă una dintre fotografiile care se potriveau cu prima imagine. Mâna stângă a pacientului

(controlată de emisfera lui dreaptă) a indicat fotografia unei lopeţi, în vreme ce mâna dreaptă a indicat imaginea unui pui. Până acum, toate bune – cele două emisfere acţionau independent şi logic. „De ce ai ales asta?”, l-a întrebat Gazzaniga pe pacient, care a răspuns verbal (controlat de centrul vorbirii din emisfera stângă): „Gheara de pui se p. otriveşte în mod evident cu puiul”. Dar apoi pacientul a privit în jos şi, observând că mâna lui stângă arăta lopata, a explicat imediat acest fapt (din nou centrul vorbirii din emisfera stângă): „Şi ai nevoie de o lopată ca să cureţi mizeria de la pui”.

Este o născocire. Emisfera dreaptă (care controlează braţul şi mâna stângă) indică în mod corect lopata, dar, cum emisfera stângă (care controlează răspunsul verbal) nu este conştientă de zăpadă, născoceşte o explicaţie, fără a fi conştientă că povesteşte. Îşi asumă responsabilitatea pentru o acţiune pe care nu a decis-o şi nu a întreprins-o niciodată, dar crede că a făcut-o.

Asta implică faptul că fiecare dintre cele două emisfere ale unui pacient cu scindare cerebrală îşi are propria conştienţă. Emisferele para nu fi conştiente că trupul lor este controlat, efectiv, de două creiere, deoarece acestea învaţă să se coordoneze între ele, iar deciziile lor sunt suficient de bine aliniate şi consecvente pentru ca fiecare să creadă că deciziile celeilalte îi aparţin.

Experimentul lui Gazzaniga nu dovedeşte că un individ normal, cu un corp calos funcţional, are două jumătăţi de creier conştiente, dar este sugestiv în privinţa acestei posibilităţi. Corpul calos permite o colaborare eficientă între cele două jumătăţi de creier, dar asta nu înseamnă neapărat că nu sunt nişte minţi separate. Fiecare poate fi păcălită şi făcută să creadă că a luat toate deciziile, deoarece acestea ar fi suficient de apropiate faţă de ce ar decide fiecare în mod independent şi, la urma urmelor, fiecare are suficient de multă influenţă asupra oricărei decizii (graţie coroborării cu cealaltă emisferă prin intermediul corpului calos). Astfel, fiecăreia dintre cele

două minţi i s-ar părea că deţine controlul.

Cum ai testa situaţia în care ambele sunt conştiente? Ele pot fi evaluate sub aspectul corelării neurologice a conştienţei, adică exact ce a făcut Gazzaniga. Experimentele acestuia arată că fiecare emisferă acţionează ca un creier separat Povestirea nu este limitată la emisferele creierului; fiecare dintre noi facem aşa ceva în mod curent. Fiecare emisferă este aproape la fel de inteligentă ca omul, deci, dacă credem că creierul omului este conştient, atunci trebuie să concluzionăm că fiecare emisferă este conştientă în mod independent. Putem evalua corelările neurologice şi putem realiza propriile noastre experimente mintale (de exemplu, considerând că, dacă două emisfere cerebrale fără un corp calos funcţional constituie două minţi conştiente separate, acelaşi lucru e valabil şi pentru două emisfere cu conexiune funcţională între ele), dar orice încercare de detectare mai directă a conştienţei fiecărei emisfere ne pune din nou faţă în faţă cu lipsa unui test ştiinţific pentru conştientă. Iar dacă acceptăm că fiecare emisferă a creierului este conştientă, atunci încuviinţăm şi că aşa-numita activitate inconştientă de la nivelul neocortexului (care constituie cea mai mare parte a activităţii sale) are de asemenea o conştienţă independentă? Sau poate că are mai multe, nu doar una. Pe bună dreptate, Marvin Minsky se referă la creier ca la o „societate a minţii”.13

În cadrul unui alt experiment pe un creier scindat, cercetătorii au arătat cuvântul „clopot” creierului drept şi cuvântul „muzică” creierului stâng. Pacientul a fost întrebat ce cuvânt vede. Centrul vorbirii controlat de emisfera stângă a spus „muzică”. Subiectului i-a fost prezentat apoi un grup de fotografii şi i s-a cerut să indice imaginea cel mai strâns legată de cuvântul care abia îi fusese arătat. Braţul controlat din emisfera dreaptă a arătat clopotul. Când a fost întrebat de ce a indicat clopotul, centrul vorbirii care controlează vorbirea, situat în emisfera stângă, a replicat: „Ei bine, muzica, ultima

oară când am auzit o muzică a fost când au bătut clopotele aici, afară". A dat această explicaţie deşi mai existau şi alte fotografii dintre care putea să aleagă, mult mai strâns legate de muzică.

Din nou, este vorba de o născocire. Emisfera stângă explică de parcă ar fi decizia ei, pe care însă nu a luat-o niciodată şi nici nu a pus-o vreodată în aplicare. Nu face acest lucru pentru a-şi acoperi un prieten (respectiv, cealaltă emisferă), ci crede realmente că decizia i-a aparţinut.

Aceste reacţii şi decizii se pot extinde şi la răspunsurile emoţionale. Un pacient adolescent cu creierul scindat – astfel încât ambele emisfere auzeau – a fost întrebat „Cine este... preferată" şi pe urmă i s-a servit cuvântul „iubita" doar pentru emisfera dreaptă, prin urechea stângă. Gazzaniga relatează că subiectul a roşit şi a avut un comportament jenat, o reacţie adecvată pentru un adolescent care este întrebat despre iubita lui. Însă centrul vorbirii din emisfera stângă a relatat că nu auzise niciun cuvânt şi a cerut o clarificare: „Cine anume e preferată?" Când i s-a cerut din nou să răspundă la întrebare, de această dată în scris, mâna stângă, controlată de emisfera dreaptă, a scris numele iubitei sale.

Testele lui Gazzaniga nu sunt experimente mintale, ci veritabile experimente ale minţii. Deşi oferă o perspectivă interesantă asupra subconştientului, ele vorbesc şi mai direct despre liberul arbitru. În fiecare dintre aceste cazuri, una dintre emisfere crede că a luat o decizie pe care, în realitate, n-a luat-o. În ce măsură este adevărat acest lucru pentru deciziile pe care le luăm în fiecare zi?

Să vedem cazul unei paciente, o fetiţă de zece ani bolnavă de epilepsie. Neurochirurgul Itzhak Fried a realizat o intervenţie chirurgicală cu pacienta trează (ceea ce este fezabil, deoarece la nivelul creierului nu există receptori ai durerii).14 De câte Ori el stimula un anumit punct de pe neocortex, ea începea să râdă. La început, echipa de chirurgi s-a gândit că era posibil să declanşeze

un fel de reflex al râsului, dar şi-a dat repede seama că stimula, practic, percepţia umorului. Părea că descoperiseră un punct de pe neocortex – în mod evident există mai multe, nu doar unul – care recunoaşte percepţia umorului. Fetiţa nu doar că râdea-situaţia chiar i se părea comică, deşi nu se schimbase nimic în stare ei în afara faptului că îi era stimulat acel punct al neocortexului. Când medicii au întrebat-o de ce râde, nu numai că a răspuns în sensul general A nu, fără vreun motiv anume" sau „Tocmai mi-aţi stimulat creierul", ci a şi născocit imediat un motiv. A arătat ceva din încăpere şi a încercat să explice de ce era amuzant lucrul respectiv. „Sunteţi atât de amuzanţi, cum staţi aşa, acolo", a fost comentariul ei tipic.

Aparent, suntem foarte dornici să ne explicăm şi să ne raţionalizăm acţiunile, chiar şi când nu noi suntem autorii deciziilor care au condus la ele. Aşadar, cât de responsabili suntem de deciziile noastre? Să ne gândim la următoarele experimente ale lui Benjamin Libet (1916 – 2007), profesor de fiziologie la Universitatea California din Davis. Libet le-a cerut participanţilor să se aşeze în faţa unui cronometru, cu electrozi de electroencefalograme ataşaţi pe cap. l-a instruit să efectueze sarcini simple, cum ar fi să apese un buton sau să mişte un deget. Participanţilor li s-a cerut să noteze timpul de pe cronometru „prima oară când deveneau conştienţi de dorinţa sau imboldul de a acţiona". Testele au indicat o marjă de eroare de numai 50 de milisecunde la aceste autoevaluări realizate de subiecţi. De asemenea, au cronometrat o medie de 200 de milisecunde între momentul în care subiecţii şi-au exprimat conştientizarea nevoii presante de a acţiona şi actul propriu-zis.15

Cercetătorii au urmărit de asemenea şi semnalele EEG provenite de la creierele subiecţilor. Activitatea cerebrală implicată în iniţierea acţiunii de către cortexul motor (care este răspunzător de realizarea acţiunii) s-a manifestat practic, în medie, cu 500 de milisecunde înainte de îndeplinirea sarcinii. Asta înseamnă următorul lucru:

cortexul motor se pregătea să îndeplinească sarcina cu aproximativ o treime de secundă înainte ca subiectul să fie conştient măcar că a luat decizia de a face lucrul respectiv.

Implicaţiile experimentelor lui Libet au fost dezbătute intens. Libet a concluzionat că însăşi conştientizarea de către noi a luării unei decizii pare să fie o iluzie şi că, până la urmă, „conştienta nu intră în buclă". Filosoful Daniel Dennett a comentat: „Acţiunea este iniţial precipitată într-o anumită parte a creierului şi de acolo pornesc semnalele pentru muşchi, oprindu-se în drum să-ţi spună ţie, agentul conştient, ce se petrece (dar, ca toţi bunii oficiali, lăsându-te pe tine, preşedintele mult-prea-important, să rămâi cu iluzia că tu ai început tot)".16 În acelaşi timp, Dennett a pus sub semnul îndoielii intervalele de timp înregistrate de experiment, în esenţă susţinând că era posibil ca subiecţii să nu fi fost, de fapt, conştienţi de momentul în care au devenit conştienţi de decizia de a acţiona. Te-ai putea întreba: dacă subiectul nu este conştient când devine conştient de luarea unei decizii, atunci cine e conştient? însă aspectul fundamental este surprins foarte bine – după cum am arătat anterior, motivul pentru care suntem conştienţi e departe de a fi clar.

Vilayanur Subramanian „Rama" Ramachandran (născut în anul 1951), american de origine indiană, specialist în neuroştiinţă, explică situaţia puţin diferit. Dat fiind faptul că la nivelul neocortexului avem aproximativ 30 de miliarde de neuroni, aici se petrec permanent foarte multe lucruri, dar suntem conştienţi de foarte puţine dintre acestea. Deciziile, mari şi mici, sunt procesate continuu de neocortex, iar soluţiile propuse ies la iveală în conştienţa noastră. În locul liberului arbitru, Ramachandran sugerează că ar trebui să vorbim despre „libera opoziţie" – respectiv forţa de a respinge soluţiile propuse de părţile nonconştiente ale neocortexului.

Să ne gândim la analogia cu o campanie militară. Oficialii militari pregătesc o recomandare pentru

preşedinte. Anterior primirii aprobării de la acesta, ei realizează o activitate pregătitoare care să permită punerea deciziei în aplicare. Într-un anumit moment, decizia propusă îi este prezentată preşedintelui, care o aprobă şi apoi se întreprinde restul misiunii. Deoarece „creierul" reprezentat de această analogie implică atât procesele inconştiente de la nivelul neocortexului (respectiv, oficialii de sub preşedinte), cât şi procesele sale conştiente (preşedintele), am vedea atât activitatea neuronală, cât şi acţiunile propriu-zise ca desfăşurându-se anterior luării deciziei oficiale. Întotdeauna, într-o anumită situaţie, se poate ajunge la dezbateri legate de mărimea răgazului pe care oficialii de sub preşedinte i-l acordă acestuia pentru acceptarea sau respingerea unei recomandări şi în mod cert preşedinţii Americii au făcut şi una, şi alta. Însă nu ar trebui să ne surprindă faptul că activitatea mintală, chiar şi la nivelul cortexului motor, poate începe înainte ca noi să fim conştienţi că exista o decizie care trebuia luată.

Experimentele lui Libet subliniază că, la nivelul creierului nostru, există o activitate foarte intensă, care stă la baza deciziilor noastre şi care nu este conştientă. Ştiam deja că mare parte din ceea ce se petrece în neocortex nu este conştient; prin urmare, nu ar trebui să fie surprinzător faptul că acţiunile şi deciziile noastre îşi au originea atât în activitatea inconştientă, cât şi în cea conştientă. Este importantă această distincţie? Dacă. deciziile noastre rezultă şi din una, şi din alta, ar trebui să conteze dacă separăm partea conştientă de cea inconştientă? Nu cumva ne reprezintă creierul ambele aspecte? Nu suntem, oare, în cele din urmă, responsabili pentru tot ceea ce se petrece în creierul nostru? „Da, am tras în victimă, dar nu sunt responsabil, pentru că nu eram atent" probabil că este o apărare slabă. Deşi există unele temeiuri juridice înguste pe baza cărora o persoană nu poate fi făcută răspunzătoare pentru deciziile sale, în general suntem făcuţi răspunzători pentru toate alegerile

pe care le facem.

Observaţiile şi experimentele pe care le-am citat mai înainte constituie experimente mintale privitoare la liberul arbitru, un subiect care, asemenea celui al conştienţei, a fost dezbătut încă de la Platon încoace. Însăşi sintagma „liber arbitru” datează încă din secolul al treisprezecelea, dar ce anume înseamnă ea mai exact?

Dicţionarul Merriam-Webstero defineşte ca fiind „libertatea omului de a face alegeri care nu sunt determinate de cauze anterioare sau de intervenţia divină”. Vei observa, cititorule, că această definiţie este fără speranţă de circulară: „Liberul arbitru este libertatea…” Lăsând deoparte ideea unei intervenţii divine care ar sta împotriva liberului arbitru, există un element util al acestei definiţii – şi anume ideea că decizia „nu este determinată de cauze anterioare”. Voi reveniimediat la asta.

Enciclopedia Stanford de filosofie afirmă că liberul arbitru este „capacitatea agenţilor raţionali de a alege un curs al acţiunii dintre mai multe alternative”. Potrivit acestei definiţii, un simplu computer este capabil de liber arbitru, astfel încât e mai puţin util decât definiţia de dicţionar.

Wikipedia este, de fapt, ceva mai bună. Defineşte liberul arbitru ca fiind „abilitatea agenţilor de a face alegeri libere de anumite genuri de constrângeri… Constrângerea care a generat preocuparea dominantă a fost… determinismul”. Din nou se foloseşte cuvântul „liber” pentru a defini liberul arbitru, dar nu se lămureşte ceea ce a fost considerat ca fiind principalul duşman al liberului arbitru: determinismul. Din acest punct de vedere, definiţia din Merriam-Webster de mai înainte este practic similară referinţei la deciziile care „nu sunt determinate de cauze anterioare”.

Aşadar, ce vrem să spunem prin determinism? Dacă introduc „2 + 2” într-un calculator şi acesta afişează rezultatul „4”, pot spune că acest calculator şi-a afişat

liberul arbitru decizând să afişeze acel „4”? Nimeni nu ar accepta asta ca demonstrare a liberului arbitru, deoarece „decizia” a fost predeterminată de mecanismele interne ale calculatorului şi de imput. Dacă introduc un calcul mai complicat, ajung la aceeaşi concluzie în privinţa lipsei de liber arbitru.

Dar Watson atunci când dă răspunsurile la întrebările de la *JeopardyW* Cu toate că deliberările lui sunt mult mai complexe decât cele ale calculatorului, foarte puţini observatori ar atribui deciziile sale liberului arbitru. Niciun om nu ştie cum funcţionează programele lui, dar putem identifica un grup de oameni capabili să-i descrie toate metodele. Mai important, outputul lui Watson este determinat de (1) toate programele lui, în momentul în care este formulată întrebarea; (2) întrebarea propriu-zisă; (3) starea parametrilor interni care îi influenţează deciziile şi (4) miile de miliarde de bytes de baze de cunoştinţe, inclusiv enciclopedii. Outputul lui este determinat pe baza acestor patru categorii de informaţii. Am putea să speculăm că adresarea aceleiaşi întrebări ar aduce întotdeauna acelaşi răspuns, dar Watson e programat să înveţe din experienţele sale, astfel încât există posibilitatea ca răspunsurile ulterioare să fie diferite. Pe de altă parte, asta nu contrazice analiza pe care o fac acum; dimpotrivă, constituie doar o schimbare la punctul 3, parametrul care îi controlează lui Watson deciziile.

Aşadar, prin ce se deosebeşte, mai exact, un om de Watson, astfel încât omului să-i putem atribui liber arbitru, pe când programului de computer nu? Cu toate că Watson este un jucător mai bun de *Jeopardy*! comparativ cu majoritatea oamenilor, nu e nici pe departe la fel de complex ca un neocortex uman. Watson posedă, într-adevăr, foarte multe cunoştinţe şi foloseşte metode ierarhice, însă complexitatea gândirii lui ierarhice e în continuare considerabil mai redusă decât cea a omului. Deci diferenţa se referă la scara de complexitate a gândirii sale ierarhice? Se poate argumenta că problema nu se

reduce la asta. În discuţia mea referitoare la subiectul conştienţei, am menţionat că propriul meu salt de gândire rezidă în faptul de a considera că un computer care a trecut Testul Turing este conştient. Nici cele mai bune *chatbot-m* nu sunt capabile, astăzi, să facă acest lucru (deşi sunt permanent îmbunătăţite), astfel încât concluzia mea cu privire la conştienţă este o chestiune de nivel de performanţă al entităţii. Poate că acelaşi lucru este adevărat şi în legătură cu faptul că îi atribui liber arbitru.

Conştienţă reprezintă, într-adevăr, o diferenţă filosofică între creierul omului şi programele software contemporane. Considerăm creierul omului ca fiind conştient, dar nu atribuim – încă – aceeaşi calitate şi programelor software. Să fie acesta factorul pe care-l căutăm şi care stă la baza liberului arbitru?

Un simplu experiment mintal ar sprijini ideea că, într-adevăr, conştienţă este o parte vitală a liberului arbitru. Să ne gândim la o situaţie în care cineva îndeplineşte o acţiune fără să-şi dea seama ce face – acţiunea este realizată exclusiv prin activitatea nonconştientă a creierului persoanei respective. Am considera oare că aceasta e o manifestare a liberului arbitru? Majoritatea oamenilor ar răspunde *nu*. Dacă acţiunea ar fi nocivă, probabil că am considera chiar persoana respectivă răspunzătoare, dar am căuta acţiuni conştiente de dată recentă care ar fi putut-o determina să întreprindă acţiuni fără a fi conştientă – ca de pildă să bea prea mult sau pur şi simplu să nu se pregătească adecvat pentru a-şi cântări deciziile înainte de a le pune în aplicare.

Potrivit unor comentatori, experimentele lui Libet sunt împotriva liberului arbitru, scoţând în evidenţă măsura în care procesul de luare a deciziilor nu e conştient. De vreme ce în rândul filosofilor există un consens rezonabil în privinţa faptului că liberul arbitru presupune şi luarea conştientă a deciziilor, aceasta pare să fie o cerinţă esenţială a liberului arbitru. Pe de altă parte,

totuşi, pentru mulţi observatori, conştienţa este o condiţie necesară, dar nu suficientă. Dacă deciziile noastre – conştiente sau nu – sunt predeterminate înainte să le luăm, cum putem spune că deciziile noastre sunt libere? Această poziţie, care susţine că liberul arbitru şi determinismul nu sunt compatibile, este cunoscută sub denumirea de incompatibilism. De exemplu, filosoful american Carl Ginet (născut în anul 1932) argumentează că, dacă evenimentele din trecut, prezent şi viitor sunt determinate, atunci se poate considera că nu avem niciun fel de control asupra lor sau asupra consecinţelor lor. Deciziile şi acţiunile noastre aparente sunt pur şi simplu parte a acestei succesiuni predeterminate. Pentru Ginet, asta elimină liberul arbitru.

Însă nu toată lumea consideră că determinismul este compatibil cu conceptul de liber arbitru. Compatibiliştii susţin, în esenţă, că eşti liber să decizi ce vrei, chiar dacă ceea ce decizi este sau poate fi determinat. Daniel Dennett, de exemplu, susţine cu argumente că, în vreme ce viitorul poate fi determinat din starea prezentului, lumea este atât de încâlcit complexă, încât de fapt nu avem cum şti ce ne va aduce viitorul. Putem identifica trăsăturile pe care Dennett le numeşte „aşteptări” şi suntem pe bună dreptate liberi să ducem la îndeplinire acţiuni care nu sunt în acord cu aceste aşteptări. Ar trebui să luăm în considerare felul cum arată deciziile şi acţiunile noastre comparativ cu aceste aşteptări şi nu cu viitorul nostru determinat teoretic, pe care, de fapt, nu-l putem cunoaşte. Acest lucru, susţine Dennett; este suficient pentru a confirma liberul arbitru.

Gazzaniga formulează şi o poziţie compatibilistă: „Suntem personal agenţi responsabili şi trebuie să fim consideraţi răspunzători pentru acţiunile noastre chiar dacă trăim într-o lume determinată”.17 Cinicul ar putea să interpreteze această perspectivă la modul: Nu ai control asupra acţiunilor tale, dar oricum o să dăm vina pe tine.

Unii gânditori resping ideea liberului arbitru. Filosoful scoţian David Hume (1711 – 1776) a descris-o pur

şi simplu ca fiind o chestiune „verbală” caracterizată de o „falsă senzaţie sau de o pretinsă experienţă”.18 Filosoful german Arthur Schopenhauer (1788 – 1860) a scris că „oricine se consideră a *priori* perfect liber, chiar şi în acţiunile sale individuale, şi crede că în orice moment poate să înceapă un alt gen de viaţă... Dar a *posteriori*, prin experienţă, constată spre uimirea sa că nu este liber, ci e supus necesităţii, că, în ciuda tuturor hotărârilor sale, nu-şi schimbă purtarea şi că, de la începutul vieţii sale şi până la sfârşitul ei, trebuie să poarte chiar caracterul pe care el însuşi îl condamnă”.19

Eu aş adăuga mai multe aspecte aici. Conceptul de liber arbitru – şi responsabilitate, care este o idee strâns legată – e util şi pe bună dreptate vital pentru menţinerea ordinii sociale, indiferent dacă liberul arbitru există sau nu. Aşa cum conştienţa există clar ca memă, tot aşa există şi liberul arbitru. Tentativele de dovedire a existenţei sale sau măcar de definire a ei pot deveni disperant de circulare, dar aproape toată lumea crede în această idee. Părţi substanţiale ale neocortexului nostru de nivel superior sunt dedicate convingerii că facem alegeri libere şi că suntem răspunzători de acţiunile noastre. Într-un sens strict filosofic, acest lucru este adevărat sau măcar posibil, iar dacă că nu am avea asemenea credinţe, societatea ar fi cu mult mai rea.

Mai mult de-atât, lumea nu este în mod necesar determinată. Am discutat mai înainte două perspective asupra mecanicii cuantice care se deosebesc din punctul de vedere al relaţiei câmpurilor cuantice cu un observator. O interpretare populară a perspectivei observatorului conferă un rol conştienţei: particulele nu-şi lămuresc ambiguitatea cuantică până în momentul în care sunt observate de un observator conştient. Mai există o scindare în filosofia evenimentelor cuantice care ne influenţează discuţia despre liberul arbitru, scindare care evoluează în jurul întrebării: Evenimentele cuantice sunt determinate sau aleatorii?

Cea mai obişnuită interpretare a unui eveniment cuantic spune că, atunci când funcţia de undă care constituie o particulă „colapsează”, localizarea particulei devine sigură. Pentru foarte multe asemenea evenimente va exista o distribuţie predictibilă (motiv pentru care funcţia de undă este considerată a fi o distribuţie probabilistică), dar hotărârea privitoare la fiecare particulă care să sufere colapsul funcţiei de undă este aleatorie. Interpretarea în sens opus este deterministă: mai exact, există o variabilă ascunsă pe care nu suntem capabili s-o detectăm separat, dar a cărei valoare determină poziţia particulei. Valoarea sau faza variabilei ascunse în momentul în care funcţia de undă colapsează determină poziţia particulei. Cei mai mulţi fizicieni specialişti în mecanica cuantică par să agreeze ideea unei hotărâri aleatorii potrivit câmpului de probabilitate, dar ecuaţiile mecanicii cuantice permit existenţa unei asemenea variabile ascunse.

Prin urmare, s-ar putea ca lumea să nu fie deloc determinată. Potrivit interpretării mecanicii cuantice făcute prin prisma undei de probabilitate, la cel mai elementar nivel al realităţii există o sursă permanentă de incertitudine. Pe de altă parte însă, această observaţie nu rezolvă neapărat motivele de îngrijorare ale incompatibiliştilor. Este adevărat că, în această interpretare a mecanicii cuantice, lumea nu e determinată, dar conceptul nostru de liber arbitru ajunge până dincolo de deciziile şi acţiunile pur aleatorii. Cei mai mulţi dintre incompatibilişti ar considera conceptul de liber arbitru incompatibil şi cu faptul că deciziile noastre sunt în esenţă accidentale. Liberul arbitru pare să presupună o luare intenţionată a deciziilor.

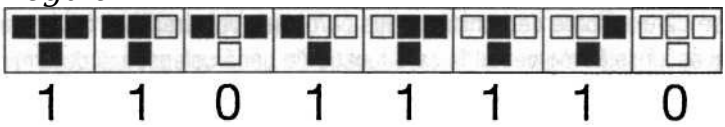
Doctorul Wolfram propune un mod de rezolvare a acestei dileme. Cartea lui, *A New Kind of Science – Un nou gen de ştiinţă*, apărută în anul 2002, prezintă o perspectivă cuprinzătoare asupra ideii de automate celulare şi asupra rolului lor în toate aspectele vieţii noastre. Automatul

celular este un mecanism în care valoarea celulelor de informaţie este recomputată continuu în funcţie de celulele din jurul ei. John von Neumann a creat o maşină teoretică autoreplicativă denumită constructor universal, care a fost, probabil, primul automat celular.

Doctorul Wolfram îşi ilustrează teza cu cel mai simplu automat celular cu putinţă, un grup de celule situate pe o linie unidimensională. În fiecare moment din timp, fiecare celulă poate să aibă două valori: neagră sau albă. Valoarea fiecărei celule este recomputată pentru fiecare ciclu. Valoarea unei celule pentru următorul ciclu este o funcţie care depinde şi de valoarea ei curentă, şi de valoarea celor doi vecini alăturaţi. Fiecare automat celular este caracterizat de o regulă care determină felul cum computăm dacă o celulă e neagră sau albă în următorul ciclu.

Să ne gândim, de exemplu, la ceea ce doctorul Wolfram numeşte regula 222.

Regula 222

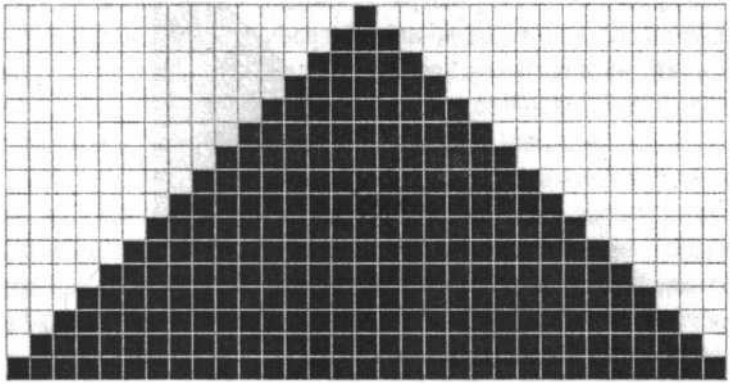


Cele opt posibile combinaţii ale valorilor celulei care este recomputată şi ale vecinilor ei din stânga şi din dreapta sunt prezentate pe rândul de sus. Noua valoare este reprezentată pe rândul de jos. Astfel, de exemplu, dacă celula este neagră şi cei doi vecini ai săi sunt de asemenea negri, celula rămâne neagră şi în următoarea generaţie (vezi subregula din extrema stângă a regulii 222). Dacă celula este albă, vecina ei din stânga e albă şi vecina din dreapta e neagră, atunci va deveni neagră în următoarea generaţie (vezi subregula regulii 222, a doua din dreapta).

Universul acestui automat celular simplu constă dintr-un singur rând de celule. Dacă începem cu o singură celulă neagră din mijloc şi arătăm evoluţia celulelor pe mai multe generaţii (în care fiecare rând, în sens descendent,

reprezintă o nouă generaţie de valori), rezultatul regulii 222 se prezintă după cum urmează:

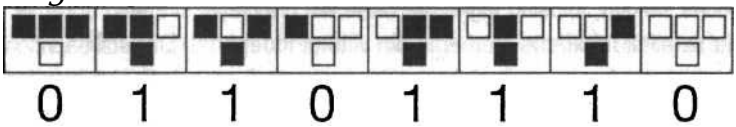
Regula 222



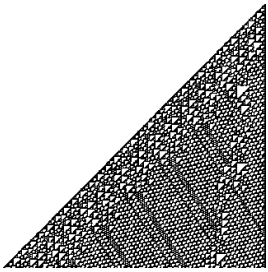
Automatul se bazează pe o regulă şi regula stabileşte dacă celula va fi neagră sau albă în funcţie de cele opt modele posibile care există în generaţia curentă. Astfel, există 28 = 256 de reguli posibile. Doctorul Wolfram a enumerat toate cele 256 astfel de automate posibile şi i-a atribuit fiecăruia un cod Wolfram de la o la 255. Interesant este că aceste 256 de maşini teoretice au proprietăţi foarte diferite. Automatul din clasa I, cum îl numeşte Wolfram, de pildă regula 222, creează modele foarte predictibile. Dacă ar fi să te întreb ce valoare va avea celula din mijloc după o mie de miliarde de mii de miliarde de repetări ale regulii 222, ai putea să răspunzi cu uşurinţă: neagră.

Şi mai interesante, pe de altă parte, sunt automatele de clasa a IV-a, ilustrate de regula 110.

Regula 110



Generaţiile multiple ale acestui automat arată astfel:



Aspectul interesant al automatului regulii 110 şi al automatelor din clasa a IV-a în general rezidă în faptul că rezultatele sunt complet impredictibile. Aceste rezultate trec şi cele mai stricte teste matematice ale caracterului aleatoriu, dar nu generează zgomot: sunt modele care se repetă, dar se repetă în moduri ciudate şi imprevizibile. Dacă te-aş întreba care este valoarea unei anumite celule după o mie de miliarde de mii de miliarde de repetări, nu ai putea să-mi răspunzi fără să treci această maşină prin toate generaţiile respective. Soluţia este clar determinată, deoarece avem de-a face cu o maşină deterministă foarte simplă, dar e complet impredictibilă în absenţa rulării propriu-zise a maşinii.

Teza primară a doctorului Wolfram este că lumea e un mare automat celular din clasa a IV-a. Motivul pentru care cartea lui se intitulează *A New KM of Science – Un nou fel de ştiinţă* este acela că teza vine în opoziţie cu majoritatea celorlalte legi ştiinţifice. Dacă am avea un satelit care ar orbita în jurul Pământului, am putea anticipa locul unde se va afla peste cinci ani de-acum înainte fără să trebuiască să trecem prin fiecare moment al unui proces simulat, folosind legile relevante ale gravitaţiei şi aflând unde se va afla satelitul în diferite momente din viitorul îndepărtat. Dar starea viitoare a automatului celular din clasa a IV-a nu poate fi anticipată fără simularea fiecărui pas. Dacă Universul este un gigantic automat celular, cum postulează doctorul Wolfram, nu ar putea să existe un computer suficient de mare – pentru că fiecare computer ar fi un subset al Universului – pentru a putea să efectueze o asemenea

simulare. Prin urmare, starea viitoare a Universului este complet imposibil de cunoscut, cu toate că e deterministă.

Astfel, cu toate că sunt determinate (deoarece corpul şi creierul nostru sunt parte a unui univers determinist), deciziile noastre sunt, totuşi, în mod inerent imprevizibile, deoarece trăim într-un automat de clasa a IV-a (şi suntem parte a lui). Nu putem anticipa viitorul unui automat de clasa a IV-a decât în cazul în care lăsăm acest viitor să se desfăşoare. Pentru doctorul Wolfram, este un fapt suficient pentru acceptarea liberului arbitru.

Nu trebuie să ne uităm la Univers pentru a vedea evenimente viitoare care sunt determinate dar imprevizibile. Niciunul dintre oamenii de ştiinţă care au lucrat la Watson nu poate să anticipeze ce va face acesta, deoarece programul este pur şi simplu prea variat şi prea complex, iar performanţa lui se bazează pe cunoştinţe mult prea extinse pentru a putea fi gestionate de om. Dacă credem că omul manifestă liber, arbitru, va trebui să acceptăm că viitoarele versiuni ale lui Watson sau maşinile asemănătoare acestuia îl pot manifesta şi ele.

Propriul meu salt de credinţă este acela că omul are liber arbitru şi, în timp ce acţionez ca şi cum aşa este, sunt obligat să găsesc exemple din propriile mele decizii care să ilustreze asta. Să ne gândim la decizia de a scrie această carte – nu eu am luat-o. Dimpotrivă, ideea cărţii a luat decizia în locul meu. În general, mă văd captivul ideilor care par să se implanteze în neocortexul meu şi să preia comanda. Cum rămâne cu decizia de a mă căsători, pe care am luat-o (în colaborare cu cealaltă persoană) în urmă cu treizeci şi şase de ani? La momentul respectiv, am urmat programul obişnuit de a fi atras – şi de a urmări – o fată frumoasă. Pe urmă m-am îndrăgostit. Unde e liberul arbitru aici?

Dar cum rămâne cu micile decizii pe care le iau zi de zi – de exemplu, cuvintele pe care aleg să le scriu în cartea mea? Pornesc de la o foaie albă de hârtie. Nimeni nu-mi spune ce să fac. Nu există niciun redactor care să mă

urmărească peste umăr. Alegerile mele depind *exclusiv* de mine. Sunt liber – *total* liber să scriu *orice*…

Ah, *grok*…

Grok? în regulă, am comis-o – mi-am folosit în sfârşit liberul arbitru. Voiam să scriu cuvântul „vreau”, dar am luat decizia de a scrie ceva total neaşteptat. Probabil că este prima oară când am reuşit să-mi exercit liberul arbitru pur.

Sau nu.

(*Kr* trebui să fie. evident că aceasta nu a fost o manifestare a liberului arbitru ci, dimpotrivă, o încercare de a ilustra o idee (şi, probabil, un simţ al umorului slăbuţ).

Deşi împărtăşesc convingerea lui Descartes că sunt conştient, nu sunt atât de convins în privinţa liberului arbitru. Este dificil să eviţi concluzia lui Schopenhauer că „poţi să faci ce vrei, dar, în orice moment al vieţii, poţi să-ţi doreşti un singur lucru şi numai unul şi nimic altceva în afara acelui lucru”.20 în orice caz, voi continua să mă port ca şi cum aş avea liber arbitru şi să cred în el, atâta vreme cât nu trebuie să explic de ce.

Identitatea

Un filosof a avut cândva următorul vis. Mai întâi i s-a arătat Aristotel şi filosoful i-a spus: „Poţi să-mi faci în cincisprezece minute o schiţă a întregii tale filosofii?” Spre surpriza filosofului, Aristotel i-a făcut o expunere excelentă, în care a comprimat un volum impresionant de material în numai cincisprezece minute. Dar, pe urmă, filosoful a avut o anumită obiecţie la care Aristotel nu a putut să răspundă. Pus în încurcătură, Aristotel a dispărut.

Pe urmă, i-a apărut Platon. S-a întâmplat iar acelaşi lucru, iar obiecţia filosofului la Platon a fost aceeaşi ca la Aristotel. Nici Platon nu a putut răspunde şi a dispărut.

Pe urmă i s-au arătat toţi filosofii celebri din istorie, unul câte unul, şi filosoful nostru i-a alungat pe toţi, pe rând, cu aceeaşi obiecţie. După ce a dispărut şi ultimul filosof, omul nostru şi-a spus: „Ştiu că dorm şi că visez

toate aceste lucruri. Dar am descoperit respingerea universală pentru toate sistemele filosofice! Mâine dimineaţă, după ce mă voi trezi, probabil că o s-o uit, iar lumea o să piardă într-adevăr ceva!” Cu o voinţă de fier, filosoful s-a forţat să se trezească, s-a repezit la masă şi şi-a aşternut pe hârtie respingerea universală. Pe urmă a sărit înapoi în pat şi a răsuflat uşurat.

A doua zi dimineaţă, când s-a trezit, s-a dus la masă să vadă ce scrisese. Cuvintele erau: „Asta e părerea ta”.

— Raymond Smullyan, citat de David Chalmers2

Ce mă întreb. mai des decât dacă sunt conştient sau nu şi dacă îmi exercit sau nu liberul arbitru este de ce sunt conştient de experienţele şi deciziile acestei persoane anume care scrie cărţi, este pasionată de drumeţii şi plimbări cu bicicleta, ia suplimente nutritive şi aşa mai departe. Un răspuns evident ar fi „Pentru că ăsta eşti”.

Probabil că acest schimb nu este mai tautologic decât răspunsurile mele de mai sus la întrebările legate de conştiinţă şi de liberul arbitru. Dar, de fapt, am un răspuns mai bun la întrebarea de ce conştienţa mea este asociată cu această persoană anume: Pentru că acesta m-am creat să fiu.

Un aforism cunoscut spune: „Eşti ceea ce mănânci”. Şi mai adevărat este să spui: „Eşti ceea ce gândeşti”. Aşa cum am discutat, toate structurile ierarhice din neocortexul meu care îmi definesc personalitatea, abilităţile şi cunoştinţele sunt rezultatul propriilor mele gânduri şi experienţe.

Oamenii cu care aleg să interacţionez, ideile şi proiectele în care aleg să mă angajez sunt, în totalitatea lor, determinanţii primari ai persoanei care am devenit. Din acest punct de vedere, ceea ce mănânc reflectă de asemenea deciziile luate de neocortexul meu. Acceptând pentru moment latura pozitivă a dualităţii liberului arbitru, propriile mele decizii s-au concretizat în cel care sunt.

Indiferent de felul cum am ajuns să fim cei care suntem, fiecare dintre noi are dorinţa ca identitatea lui să

se menţină. Dacă nu aveai dorinţa de a supravieţui, nu te-ai afla unde eşti şi nu ai citi această carte. Orice fiinţă are acest scop – este principalul determinant al evoluţiei. Subiectul identităţii e, probabil, şi mai greu de definit decât acela al conştienţei sau al liberului arbitru, dar este – demonstrabil – cel mai important. La urma urmelor, trebuie să ştim ce suntem dacă vrem să încercăm să ne conservăm existenţa.

Gândeşte-te la acest experiment mintal: te afli cândva în viitor şi ai la dispoziţie tehnologii mai avansate decât cele de azi. Când dormi, un anumit grup îţi scanează creierul şi copiază toate detaliile ieşite în evidenţă. Echipa poate face acest lucru cu maşini de scanat de mărimea elementelor figurate ale sângelui, maşini care circulă prin capilarele din creierul tău, sau prin intermediul unei alte tehnologii neinvazive, însă are toată informaţia din creierul tău la un moment dat. De asemenea, citeşte şi înregistrează orice detaliu fizic al corpului tău care ar putea să se reflecte în starea ta de spirit, ca de pildă sistemul endocrin. Imortalizează acest „fişier at minţii” într-un corp nonbiologic care seamănă, se mişcă la fel ca tine şi are subtilitatea şi supleţea necesare pentru a trece drept tine. Dimineaţa, eşti informat asupra acestui transfer şi priveşti (poate fără să fii observat) clona minţii tale, pe care o vom denumit Tu 2. Acest Tu 2 vorbeşte despre viaţa lui ca şi cum ar fi tu şi dezvăluie cum a descoperit chiar în acea dimineaţă că a primit o nouă versiune mai durabilă a corpului, 2.O. „Hei, îmi şi place acest nou corp!”, afirmă Tu 2.

Prima întrebare la care trebuie să te gândeşti este: Tu 2 e conştient? Păi, în mod evident, pare să fie. Tu 2 trece testul pe care l-am conceput anterior, în sensul că prezintă indiciile subtile că ar fi o persoană înzestrată cu sensibilitate şi conştienţă. Dacă tu eşti conştient, înseamnă că şi Tu 2 este conştient.

Prin urmare, dacă ai dispărea, să spunem, nimeni n-ar băga de seamă. Tu 2 ar putea să se plimbe de colo-colo

pretinzând că este tu. Toţi prietenii tăi şi persoanele iubite de tine s-ar mulţumi cu această situaţie şi probabil că ar fi încântate că de-acum ai un corp şi un substrat mintal mai durabile decât înainte. Poate că prietenii tăi mai înclinaţi către filosofie şi-ar exprima îngrijorarea, dar, în cea mai mare parte, toată lumea ar fi fericită, inclusiv tu – sau cel puţin persoana care afirmă în mod convingător că ar fi tu.

Deci nu mai avem nevoie nici de vechiul corp şi nici de vechiul creier, nu? E-n regulă dacă ne descotorosim de ele?

Probabil că nu vei dori să accepţi aşa ceva. Am spus că scanarea a fost neinvazivă, aşa încât tu eşti în continuare prezent şi conştient. Mai mult de-atât, simţul identităţii continuă să te însoţească, nu s-a dus cu Tu 2, deşi Tu 2 crede că e o continuare a ta. Este posibil chiar ca Tu 2 să nici nu fie conştient că exişti sau că ai existat vreodată. De fapt, nu ai fi fost nici tu conştient de existenţa lui Tu 2 dacă nu ţi-am fi spus noi.

Concluzia noastră? Tu 2 este conştient, dar este o altă persoană, distinctă de tine – Tu 2 are o identitate diferită. Îţi este extrem de similar, cu mult mai mult decât o simplă clonă genetică, pentru că îţi împărtăşeşte inclusiv toate modelele şi conexiunile neocorticale. Sau poate că ar trebui să spun că a împărtăşit aceste modele în momentul în care a fost creat. Din acel moment, voi doi aţi pornit fiecare pe calea lui din punct de vedere neocortical. Tu eşti în continuare prezent. Nu ai aceleaşi experienţe ca Tu 2. Concluzia: Tu 2 nu este tu.

În regulă, toate bune şi frumoase până aici. Acum, să ne gândim la un alt experiment mintal – unul care este, cred eu, mai realist în termeni de surprize ale viitorului. Eşti supus unei proceduri de înlocuire a unei foarte mici porţiuni din creierul tău cu o unitate nonbiologică. Eşti convins că nu e ceva nociv şi că au fost comunicate numeroase beneficii legate de această operaţie.

Nu e ceva extrem de deplasat, deoarece această procedură le este aplicată în mod curent persoanelor care

suferă de dizabilităţi neurologice şi senzoriale, cum ar fi implanturile neuronale pentru boala Parkinson şi implanturile cohleare pentru surzi. În aceste cazuri, dispozitivul computerizat este plasat în interiorul corpului tău, însă în afara creierului, dar e conectat la creier (sau, în cazul implanturilor cohleare, la nervul auditiv). Din punctul meu de vedere, faptul că acest computer este plasat, practic, în afara creierului propriu-zis nu reprezintă ceva semnificativ sub aspect filosofic: efectiv amplificăm capacităţile creierului şi înlocuim cu un dispozitiv computerizat acele funcţii ale sale care nu mai sunt adecvate. În anii ’2030, când dispozitivele computerizate inteligente vor avea mărimea elementelor figurate ale sângelui (şi reţine că globulele albe sunt suficient de inteligente ca să recunoască şi să combată agenţii patogeni), ele vor fi introduse non-invaziv, fără să fie necesară o intervenţie chirurgicală.

Revenind la acest scenariu din viitor, ţi s-a aplicat procedura şi, cum ţi s-a promis, ea dă rezultate bune – anumite capacităţi ale tale s-au îmbunătăţit. (Ai o memorie mai bună, poate.) Deci mai eşti tot tu sau nu? Prietenii tăi cu siguranţă cred că eşti. Şi tu crezi la fel. Nu există niciun argument solid că ai fi devenit brusc o altă persoană. În mod evident, ai fost supus procedurii de efectuare a unui anumit schimb, dar eşti în continuare tu, cel dintotdeauna. Identitatea ta nu s-a modificat. Nu se pune problema că o conştienţă a altcuiva ar fi pus brusc stăpânire pe corpul tău.

În regulă. Aşadar, încurajat de aceste rezultate, decizi acum să mai urmezi o procedură, de această dată interesând o altă regiune a creierului tău. Rezultatul este acelaşi: resimţi o anumită ameliorare sub aspectul capacităţii, dar eşti în continuare tot tu.

Ar trebui să fie evident unde bat cu asta. Continui să optezi pentru alte şi alte proceduri, iar încrederea ta în acest proces nu face decât să crească, până când, în cele din urmă, ţi-ai schimbat până şi ultima parte din creier. De

fiecare dată, procedura a fost realizată cu cea mai mare grijă, pentru păstrarea tuturor modelelor şi conexiunilor tale neocorticale, astfel încât nu ţi-ai pierdut nimic din personalitate, abilităţi şi memorie. Nu a existat niciodată un Tu şi un Tu 2; ai fost doar Tu. Nimeni nu observă că încetezi să mai exişti – nici măcar tu însuţi. Şi pe bună dreptate – uită-te la tine!

Concluzia noastră: Tu continui să exişti. Nu este loc pentru nicio dilemă aici. Totul este în regulă.

Cu următoarea excepţie: după procesul treptat de înlocuire, eşti pe deplin echivalent cu Tu 2 din anteriorul experiment mintal (pe care îl voi numi *scenariu de scanare şi imortalizare*). Tu, cel de după scenariul de înlocuire, ai toate modelele şi conexiunile neocorticale pe care le aveai iniţial, doar că într-un substrat nonbiologic, ceea ce este adevărat şi pentru Tu 2 din scenariul scanării şi imortalizării. Tu, cel de după aplicarea scenariului înlocuirii treptate; ai unele capacităţi suplimentare şi o mai mare durabilitate decât aveai înaintea procesului, însă acest lucru e la fel de adevărat şi pentru Tu 2 din procesul de scanare şi imortalizare.

Dar am concluzionat că Tu 2 nu este una şi aceeaşi persoană cu tine. Iar dacă tu, după procesul de înlocuire treptată, eşti absolut echivalent cu Tu 2 de după procesul de scanare şi imortalizare, înseamnă că tu cel de după procesul de înlocuire treptată trebuie să nu mai fi tot tu.

Situaţia aceasta ne contrazice însă concluzia anterioară. Procesul de înlocuire treptată constă în mai mulţi paşi. Fiecare din aceşti paşi a părut să conserve identitatea, la fel cum am concluziona astăzi că o persoană care suferă de boala Parkinson are aceeaşi identitate şi după instalarea unui implant neuronal.22

Exact acest gen de dilemă filosofică îi determină pe unii să concluzioneze că scenariile de înlocuire nu vor avea loc niciodată (cu toate că ele se petrec deja). Dar să ne gândim la următorul lucru: în mod natural, suntem supuşi unui proces de înlocuire treptată pe parcursul întregii

noastre vieţi. Cele mai multe celule din corpul nostru sunt înlocuite continuu. (Tocmai ţi s-au înlocuit 100 de milioane de celule cât timp ai citit ultima propoziţie.) Celulele mucoasei interne a intestinului subţire se schimbă în aproximativ o săptămână, la fel ca mucoasa protectoare a stomacului. Durata de viaţă a globulelor albe ale sângelui ajunge de la câteva zile la câteva luni, în funcţie de tip. Plachetele sanguine trăiesc în jur de nouă zile.

Neuronii persistă, însă organitele şi moleculele care le compun se schimbă în decurs de o lună.23 Timpul de înjumătăţire a microtubulilor din neuron este de circa zece minute; filamentele de actină din dendrite trăiesc în jur de patruzeci de secunde; proteinele care asigură energia sinapselor sunt înlocuite la fiecare oră; receptorii NMDA din sinapse au o viaţă relativ lungă, de circa cinci zile.

Prin urmare, eşti înlocuit aproape complet în decurs de câteva luni, fapt comparabil cu scenariul înlocuirii treptate pe care l-am descris anterior. Eşti aceeaşi persoană care erai în urmă cu câteva luni? Sigur că există unele diferenţe. Poate că ai învăţat câteva lucruri noi. Dar presupui că identitatea ta se păstrează, că nu eşti distrus şi recreat continuu.

Gândeşte-te la un râu cum este cel care trece pe lângă biroul meu. Acest râu, căruia oamenii îi spun Charles, este acelaşi cu râul pe care l-am văzut ieri? Mai întâi, hai să reflectăm la ceea ce este un râu. Dicţionarul îl defineşte ca „şuvoi natural, mare, de apă curgătoare”. Potrivit acestei definiţii, râul pe care îl privesc este complet diferit de cel care era ieri. Fiecare dintre moleculele sale de apă s-a schimbat, proces care se petrece foarte repede. Filosoful grec Diogene Laertius a scris în secolul trei al erei noastre că „nu poţi să calci de două ori în apa aceluiaşi râu”.

Însă nu aşa privim în general râurile. Oamenilor le place să le privească deoarece ele sunt simboluri ale continuităţii şi stabilităţii. Potrivit concepţiei generale, râul Charles pe care l-am privit ieri este acelaşi râu pe care îl

privesc şi azi. Viaţa noastră este aproape aceeaşi. În esenţă, nu suntem materia din care este alcătuit corpul şi creierul nostru. În esenţă, aceste particule curg prin noi la fel cum curg particulele de apă prin râu. Suntem un model care se schimbă încetul cu încetul, dar are stabilitate şi continuitate, chiar dacă materia care alcătuieşte modelul se schimbă rapid.

Introducerea treptată a sistemelor nonbiologice în corpul şi creierul nostru nu va fi altceva decât un nou exemplu de înnoire continuă a părţilor care ne alcătuiesc. Asta nu ne va altera continuitatea identităţii mai mult decât înlocuirea naturală a celulelor noastre biologice. În cea mai mare parte, ne-am externalizat memoria istorică, intelectuală, socială şi personală în dispozitive electronice şi în nor. Dispozitivele cu care interacţionăm pentru accesarea acestei memorii poate că nu se află încă în interiorul corpului şi creierului nostru, dar, pe măsură ce devin din ce în ce mai mici (şi ne micşorăm tehnologia cu o rată de circa 100 în volum pe deceniu), îşi vor croi drumul către interiorul nostru, în orice caz, le vom pune într-un loc util – şi nici nu le vom pierde în acest fel. Dacă oamenii vor opta într-adevăr pentru plasarea unor dispozitive microscopice în interiorul corpului lor, acest lucru va fi în regulă, deoarece vor exista şi alte moduri de accesare a inteligenţei omniprezente a norului.

Dar să revenim la dilema pe care am propus-o mai înainte. Tu, după o perioadă de înlocuire treptată, eşti echivalent cu Tu 2 din scenariul scanării şi imortalizării, dar am decis că Tu 2 din acel scenariu nu are aceeaşi identitate ca tine. Prin urmare, cu ce ne-am ales?

Ne-am ales cu înţelegerea unei capacităţi pe care sistemele nonbiologice o au, iar sistemele biologice n-o au: capacitatea de a fi copiate pentru siguranţă, salvate şi recreate. Facem în mod curent acest lucru cu dispozitivele noastre. Când ne cumpărăm un *smartphone* nou, copiem în el toate fişierele, astfel că el are, în mare, cam aceeaşi personalitate, aceleaşi abilităţi şi amintiri ca vechiul

smartphone. Poate că ale şi 6 serie de capacităţi noi, dar conţinutul vechiului telefon ne însoţeşte mai departe. În mod similar, un program cum este Watson are în mod cert o copie de siguranţă. Dacă hardware-ul lui Watson ar fi distrus mâine, Watson ar putea fi recreat cu uşurinţă din fişierele de siguranţă stocate în nor.

Aceasta reprezintă o capacitate a lumii nonbiologice care nu există în lumea biologică. Este un avantaj, nu o limitare, motiv pentru care astăzi suntem foarte dornici să continuăm să ne încărcăm memoriile în nor. În mod cert vom continua în această direcţie, pe măsură ce sistemele nonbiologice vor căpăta tot mai multe capacităţi ale creierului nostru biologic.

Rezolvarea pe care o dau eu acestei dileme este următoarea: Nu e adevărat că Tu 2 nu eşti tot Tu – eşti Tu. Doar că acum sunt doi Tu. Nu e chiar atât de rău – dacă crezi că eşti bun, atunci doi de Tu înseamnă şi mai bine.

Ce cred că se va întâmpla de fapt este că vom continua pe calea scenariului înlocuirii treptate şi al augmentării până când, în cele din urmă, cea mai mare parte a gândirii noastre se va realiza în nor. Saltul meu de credinţă în privinţa identităţii este că aceasta se conservă prin continuitatea modelului de informaţii ce ne face oamenii care suntem. Continuitatea permite schimbarea permanentă, astfel încât, cu toate că sunt întrucâtva diferit de cel care eram ieri, am, totuşi, aceeaşi identitate. Pe de altă parte, continuitatea modelului care constituie identitatea mea nu e dependentă de substrat. Substraturile biologice sunt minunate – ne-au dus foarte departe –, însă noi creăm acum un substrat mai capabil şi mai durabil din motive cât se poate de pozitive.

Capitolul 10

LEGEA ACCELERĂRII VENITURILOR APLICATĂ CREIERULUI

Cu toate că omul ar trebui să rămână, din anumite puncte de vedere, fiinţa superioară, oare acest lucru nu este în acord cu practica naturii, care va accepta

superioritatea, sub anumite aspecte, a unor animale care au fost de mult depăşite? Oare nu le-a permis furnicii şi albinei să-şi păstreze superioritatea faţă de om prin organizarea comunităţilor lor şi prin aranjamentele sociale, păsării prin străbaterea văzduhului, peştelui prin înot, calului prin forţă şi iuţeală şi câinelui prin simţul sacrificiului de sine?

— Şamuel Butler, 1871

A fost o vreme când Pământul era, după toate aparenţele, lipsit de orice formă de viaţă animală şi vegetală şi când, potrivit opiniei celor mai buni filosofi ai noştri, era doar o sferă încinsă, cu o crustă care se răcea treptat. Dacă ar fi existat vreun om pe vremea când Pământul era în acea stare şi ar fi putut să-l vadă ca pe un fel de altă lume cu care nu avea nimic în comun, dacă în acelaşi timp ar fi fost neştiutor în ştiinţa fizicii, n-ar fi spus el, oare, că este imposibil ca fiinţe capabile să posede ceva asemănător conştienţei să evolueze din cenuşa pe care o vedea? N-ar fi negat că aceasta conţinea vreun potenţial de conştienţă? Şi totuşi, în decursul timpului, a apărut şi conştienţa. Nu este posibil, atunci, să existe chiar şi acum noi canale pentru conştienţă, deşi nu putem să le discernem, în prezent, indiciile?

— Samuel Butler, 1871

Când reflectăm la multitudinea fazelor vieţii şi conştienţei care au evoluat deja, ar însemna să ne pripim dacă am spune că nu se mai pot dezvolta şi altele şi că viaţa animală este sfârşitul tuturor lucrurilor. A fost o vreme în care focul era sfârşitul tuturor lucrurilor; o alta când piatra şi apa erau sfârşitul.

— Samuel Butler, 1871

Nu există nicio probă împotriva dezvoltării în ultimă instanţă a conştienţei în faptul că acum maşinile au prea puţină conştienţă. O moluscă nu are prea multă conştienţă. Reflectaţi la extraordinarul progres pe care l-au făcut maşinile în decursul ultimelor câteva sute de ani şi observaţi cum progresează, încetul cu încetul, regnul

animal şi vegetal. Maşinile superior organizate nu sunt atât creaturi ale zilei de ieri, cât ale ultimelor cinci minute, ca să spun aşa, raportat la trecut. Să presupunem de dragul argumentaţiei că fiinţe conştiente există de aproximativ douăzeci de milioane de ani: vedeţi ce progrese au făcut maşinile în ultima mie de ani! Nu poate lumea să mai dureze încă douăzeci de milioane de ani? Şi atunci ce anume n-ar putea ele să devină în cele din urmă?

— Samuel Butler, 1871

Teza mea fundamentală, pe care o numesc Legea accelerării veniturilor, este că mărimile fundamentale ale tehnologiei informaţiei urmează traiectorii predictibile şi exponenţiale, dezminţind prejudecata că „nu poţi prezice viitorul”. Sunt încă multe lucruri – care proiect, companie sau standard tehnic vor prevala pe piaţă, când va veni pacea în Orientul Mijlociu – care rămân de neştiut, dar raportul preţ/performanţă şi capacitatea informaţiei s-au dovedit, totuşi, remarcabil de predictibile. În mod surprinzător, aceste tendinţe rămân neperturbate de condiţii cum ar fi războiul sau pacea, prosperitatea sau recesiunea.

Un motiv principal pentru care evoluţia a creat creiere l-a reprezentat prezicerea viitorului. Când unul dintre strămoşii noştri mergea prin savană, cu mii de ani în urmă, probabil că a observat că un animal înainta către drumul pe care îl urma şi el. Putea să-şi dea seama că, dacă rămâneau pe aceleaşi direcţii, drumurile lor urmau să se intersecteze. Pe baza acestei constatări, el a decis să se îndrepte într-o altă direcţie, iar anticiparea s-a dovedit preţioasă pentru supravieţuire.

Însă aceşti predictori incluşi ai viitorului sunt liniari, nu exponenţiali, calitate care îşi are originea în organizarea liniară a neocortexului. Să ne amintim că neocortexul face în permanenţă predicţii – ce literă şi ce cuvânt vom vedea imediat, pe cine ne aşteptăm să vedem după primul colţ de drum şi aşa mai departe. Neocortexul este organizat în succesiuni liniare de paşi pentru fiecare

model, ceea ce înseamnă că gândirea exponenţială nu este ceva natural pentru noi. Şi cerebelul foloseşte predicţiile liniare. Când ne ajută să prindem o minge din zbor, face o predicţie liniară a locului în care se va afla mingea în câmpul nostru vizual şi a celui în care ar trebui să fie mâna în mănuşă, în câmpul vizual, pentru a o prinde.

După cum am subliniat, există diferenţe dramatice între progresia liniară şi cea exponenţială (patruzeci de paşi liniari înseamnă tot patruzeci, pe când exponenţiali înseamnă o mie de miliarde), ceea ce explică motivul pentru care predicţiile mele izvorâte din Legea accelerării veniturilor le par surprinzătoare, la prima vedere, multor observatori. Trebuie să ne antrenăm să gândim exponenţial. Când este vorba de tehnologia informaţiei, acesta este modul corect de gândire.

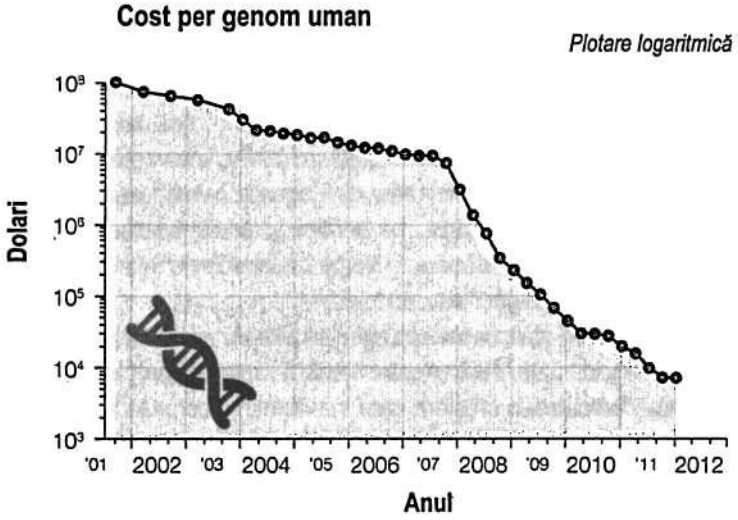
Exemplul cel mai pur pentru Legea accelerării veniturilor este creşterea perfect neaccidentată, dublu exponenţială, a raportului preţ/performanţă în domeniul computaţiei, creştere care s-a menţinut constantă timp de 110 ani, deşi lumea a cunoscut două Războaie Mondiale, Marea Recesiune şi Războiul Rece, colapsul Uniunii Sovietice, reemergenţa Chinei, recenta criză financiară şi toate celelalte evenimente notabile de secol nouăsprezece, douăzeci şi început de secol douăzeci şi unu. Unii se referă la acest fenomen ca la „Legea lui Moore”, însă ideea e greşită. Legea lui Moore – care afirmă că ia fiecare doi ani se dublează numărul de componente de pe un circuit integrat care funcţionează mai rapid pentru că şi componentele sunt mai mici – este doar una dintre multele paradigme. De fapt, creşterea exponenţială a raportului preţ/performanţă în computaţie a fost produsă de cea de-a cincea paradigmă, nu de prima.

Creşterea exponenţială a computaţiei a debutat la recensământul din 1890 din Statele Unite (primul care s-a realizat automat), folosindu-se prima paradigmă a calculului electromecanic, cu zeci de ani înainte ca Gordon Moore să se fi născut. În cartea mea *The Singularity Is*

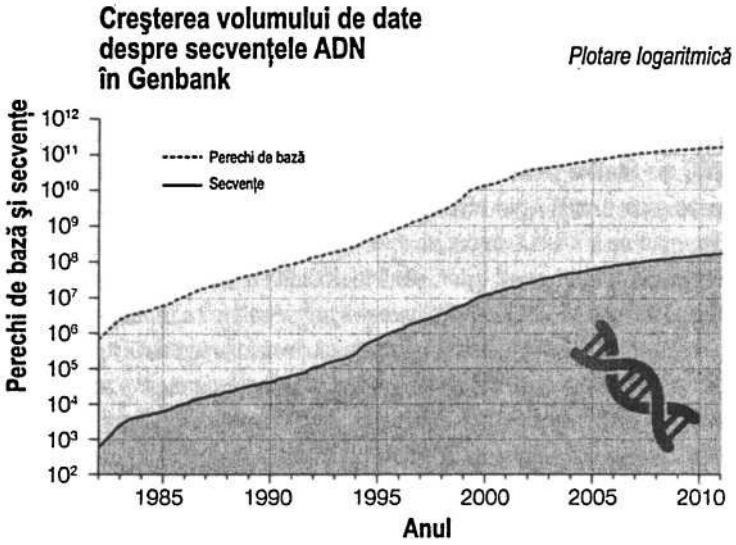
Near-Singularitatea e aproape, prezint acest grafic până în anul 2002, iar aici îl actualizez până în 2009 (vezi graficul intitulat *Creşterea exponenţială a computaţiei timp de 110 ani*). Traiectoria uşor predictibilă a continuat, în ciuda recentului recul economic.

Computaţia este cel mai important exemplu de lege a accelerării veniturilor, graţie volumului de date pe care îl avem în acest domeniu, al omniprezenţei computaţiei şi al rolului său cheie în revoluţionarea, în ultimă instanţă, a tot ceea ce ne este drag şi ne interesează. Dar nu este nici pe departe unicul exemplu. Din momentul în care tehnologia devine tehnologie a informaţiei, ea începe să se supună Legii accelerării veniturilor.

Biomedicina devine cel mai important nou domeniu al tehnologiei şi industriei care e transformat în acest fel. Progresul în medicină s-a bazat.



Costul secvenţializării unui genom de mărime umană1



Volumul de date genetice secvenţializate în lume în fiecare an2

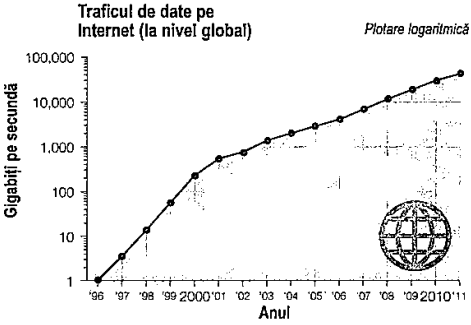
din perspectivă istorică, pe descoperiri accidentale, astfel încât avansul în cursul epocii anterioare a fost liniar, nu exponenţial. Cu toate acestea, a fost un fapt benefic: speranţa de viaţă a crescut de la douăzeci şi trei de ani în urmă cu un mileniu, la treizeci şi şapte de ani în urmă cu două secole şi la aproape optzeci de ani astăzi. Odată cu culegerea software-ului vieţii – genomul –, medicina şi biologia umană au devenit tehnologii ale informaţiei, însuşi Proiectul genomului uman a fost perfect exponenţial, volumul de date genetice dublându-se, iar costul la perechea de baze reducându-se cu jumătate în fiecare an de la înfiinţarea proiectului, în 1990.3 (Toate graficele din acest capitol au fost actualizate după publicarea cărţii mele *The Singularity Is Near-Singularitatea este aproape*.)

Actualmente, avem capacitatea de a concepe intervenţii biomedicale pe computere şi de a le testa pe simulatoare biologice, ale căror proporţii şi precizie se dublează şi ele de la an la an. Putem de asemenea să ne actualizăm propriul software desuet: interferenţa AM poate să deconecteze genele şi noi forme de terapie genetică pot să adauge gene noi nu doar unui nou-născut,

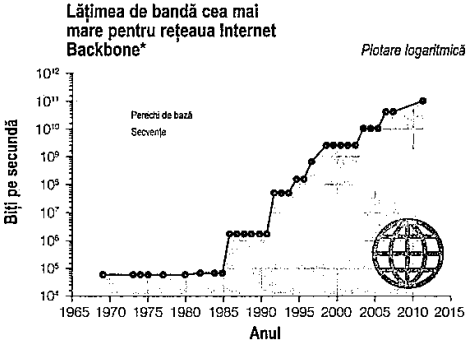
ci şi unui individ matur. Progresul tehnologiilor genetice afectează şi procesul de prelucrare a creierului prin metodele retroingineriei, prin faptul că un aspect important al său este înţelegerea modului în care genele controlează unele funcţii ale creierului, cum ar fi crearea de noi conexiuni care să reflecte cunoştinţele corticale adăugate recent. Există numeroase alte manifestări ale acestei întrepătrunderi a biologiei cu tehnologia informaţiei, pe măsură ce depăşim secvenţierea genomului în direcţia sintetizării acestuia.

O altă tehnologie a informaţiei care a cunoscut o creştere exponenţială neaccidentată o reprezintă capacitatea noastră de a comunica între noi şi de a transmite vaste baze de cunoştinţe umane. Există numeroase moduri de măsurare a acestui fenomen. Legea lui Cooper, care spune că totalul capacităţii comunicaţiilor *wireless* pe un anumit interval al spectrului radio se dublează la fiecare trei luni, a rămas valabilă din vremea în care Guglielmo Marconi a folosit telegraful fără fir pentru transmisiuni în codul Morse în anul 1897 şi până la tehnologiile comunicaţiilor 4G din zilele noastre.4 Potrivit Legii lui Cooper, volumul de informaţie care poate fi transmis pe un anumit interval al spectrului radio s-a dublat la fiecare doi ani şi jumătate, timp de mai bine de un secol. Un alt exemplu îl constituie numărul de biţi pe secundă transmişi pe Internet, care se dublează la fiecare 1, 4 ani.5

Motivul pentru care am început să fiu interesat de anticiparea anumitor aspecte ale tehnologiei este că, în urmă cu aproximativ treizeci de ani.



Lăţimea de bandă internaţională (interţări) alocată Internetului pe plan mondia16


Lățimea de bandă cea mai mare pentru rețeaua Internet Backbone*
Plotare logaritmică
Biți pe secundă
Perechi de bază
Secvențe
1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015
Anul

Cea mai mare lăţime de bandă (viteză) pentru Internet Backbone7

Internet Backbone este termenul prin care se face referire la traseele mari de date dintre reţelele mari, interconectate strategic, şi ruterele centrale de pe Internet. Aceste trasee de date sunt găzduite de centre de reţea comerciale, academice, guvernamentale, precum şi de alte centre de reţea de mare capacitate etc. (n.trad. – sursa: Wikipedia)

mi-am dat seama că soluţia pentru a avea succes ca inventator (profesie pe care am îmbrăţişat-o pe când aveam cinci ani) este sincronizarea. Cele mai multe invenţii şi cei mai mulţi inventatori au eşuat nu pentru că născocirile respective nu ar fi funcţionat, ci pentru că sincronizarea lor a fost greşită, ele apărând fie înainte de a exista toţi factorii potenţatori, fie prea târziu, ratând astfel fereastra de oportunitate.

Fiind inginer, în urmă cu aproximativ trei decenii am

început să adun date privitoare la dimensiunile tehnologiilor din diverse domenii. Când am început această activitate, nu m-am aşteptat să-mi dea o imagine clară, însă am sperat să îmi asigure o anumită orientare şi să-mi permită să fac predicţii în cunoştinţă de cauză. Scopul meu era – şi este în continuare – să-mi sincronizez eforturile din domeniul tehnologic propriu, astfel încât acestea să fie adecvate pentru lumea existentă în momentul în care îmi închei un proiect – lume care, mi-am dat seama, este foarte diferită de cea existentă în momentul de debut al proiectului.

Să ne gândim acum cât de mult şi cât de rapid s-a schimbat lumea în ultimul timp. În urmă cu numai câţiva ani, oamenii nu foloseau reţelele sociale (Facebook, de exemplu, a fost fondat în anul 2004 şi avea 901 milioane de utilizatori activi la sfârşitul lunii martie 2012) °, wiki, bloguri sau tweet. În anii 1990, cea mai mare parte din populaţia lumii nu folosea motoare de căutare sau telefoane celulare. Imaginează-ţi lumea fără ele. Pare ceva de domeniul istoriei antice, dar se întâmpla nu cu mult timp în urmă. Lumea se va schimba şi mai dramatic în viitorul apropiat.

În cursul investigaţiilor mele, am făcut o descoperire uluitoare: dacă o tehnologie este una a informaţiei, mărimile fundamentale ale raportului preţ/performanţă şi capacitate (raportată la unitatea de timp, cost sau la altă resursă) urmează traiectorii precis exponenţiale.

Aceste traiectorii depăşesc paradigmele specifice pe care se bazează (cum ar fi Legea lui Moore). Însă atunci când o paradigmă rămâne fără elan (de exemplu, momentul când inginerii nu au mai putut să reducă mărimea sau costurile tuburilor cu vid în anii 1950), se generează presiune asupra cercetării pentru crearea următoarei paradigme şi astfel debutează o nouă curbă S a progresului.

Partea exponenţială a curbei S care urmează şi vizează noua paradigmă continuă apoi curba permanent

exponenţială a măsurilor tehnologiei informaţiei. Astfel, computarea bazată pe tuburile cu vid din anii 1950 a lăsat cale liberă pentru tranzistori în anii 1960 şi pe urmă pentru circuitele integrate şi Legea lui Moore la sfârşitul anilor 1960 şi după aceea. La rândul ei, Legea lui Moore va lăsa cale liberă computării tridimensionale, ale cărei exemple timpurii sunt deja implementate. Motivul pentru care tehnologiile informaţiei sunt capabile să transcendă în mod sistematic limitele oricărei paradigme este acela că resursele necesare computaţiei, recuperării sau transmiterii unui bit de informaţie sunt din ce în ce mai mici.

Ne-am putea întreba: Chiar există limite fundamentale pentru abilitatea noastră de computaţie şi de transmitere a informaţiei indiferent de paradigmă? Răspunsul este da, pe baza felului în care înţelegem actualmente fizica computaţiei. Pe de altă parte, aceste limite nu sunt prea limitative. În ultimă instanţă, ne putem extinde inteligenţa de miliarde de ori pe baza computaţiei moleculare. După calculele mele, vom atinge aceste limite la sfârşitul acestui secol.

Este important de subliniat faptul că nu orice fenomen exponenţial reprezintă un exemplu pentru Legea accelerării veniturilor. Unii observatori interpretează greşit această lege, citând tendinţe exponenţiale care nu se bazează pe informaţie. De pildă, ei invocă faptul că aparatele de ras pentru bărbaţi au ajuns de la o lamă la două sau la patru şi apoi se întreabă unde sunt aparatele de ras cu opt lame. Numai că aparatele de ras nu înseamnă (deocamdată) tehnologie a informaţiei.

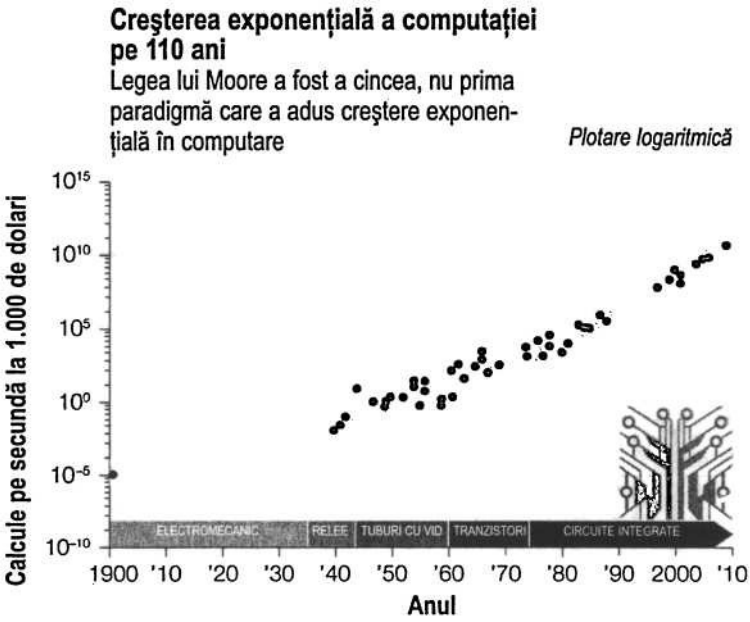
În cartea mea *Singularity Is Near-Singularitatea este aproape*, am făcut o examinare teoretică, incluzând (în anexa cărţii) o tratare matematică a motivului pentru care Legea accelerării veniturilor este remarcabil de predictibilă. În esenţă, folosim întotdeauna vârful ultimei tehnologii pentru a o crea pe următoarea. Tehnologiile se construiesc singure într-o manieră exponenţială şi acest

fenomen e uşor de măsurat dacă implică tehnologia informaţiei. În anul 1990, foloseam computere şi alte instrumente ale erei respective pentru crearea computerelor anului 1991; în anul 2012, utilizăm actualele instrumente ale informaţiei pentru a crea maşinile anilor ’2013 şi 2014. Vorbind într-un sens mai larg, această accelerare şi dezvoltare exponenţială se aplică oricărui proces în cadrul căruia evoluează modele de informaţie. Prin urmare, observăm accelerarea ritmului evoluţiei biologice, dar şi o accelerare similară (şi mult mai rapidă) a evoluţiei tehnologice, care este ea însăşi o ramificaţie a evoluţiei biologice.

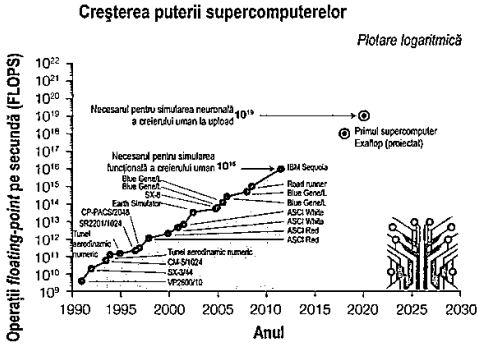
Actualmente, mă aflu în posesia unui palmares public de peste un sfert de secol de predicţii bazate pe Legea accelerării veniturilor, începând cu cele prezentate în cartea mea intitulată *The Age of Intelligent Machines – Epoca maşinilor inteligente*, pe care am scris-o la mijlocul anilor 1980. Printre exemplele de predicţii precise din această carte se numără: apariţia spre sfârşitul anilor 1990 a unei vaste reţele mondiale de comunicaţii, care va conecta oamenii din toate colţurile lumii unii cu alţii şi cu toate cunoştinţele umane; apariţia unui puternic val de democratizare generat de aceste reţele descentralizate de comunicaţii care au măturat Uniunea Sovietică; înfrângerea campionului mondial la şah înainte de 1998 şi multe altele.

Am descris pe larg Legea accelerării veniturilor, aşa cum este ea aplicată computaţiei, în *The Ageof Intelligent Machines – Epoca maşinilor inteligente*, carte în care am oferit, pentru un secol întreg, şi datele ce indică progresul dublu exponenţial al raportului preţ/performanţă al computaţiei până în anul 1998. În continuare, le completez până în anul 2009.

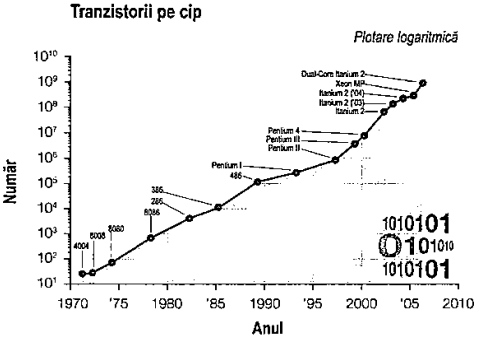
Recent, am realizat o trecere în revistă pe 146 de pagini a predicţiilor pe care le-am făcut în cărţile mele anterioare, *The Age of Intelligent Machines – Epoca maşinilor inteligente*, *The Age of Spiritual Machines –*



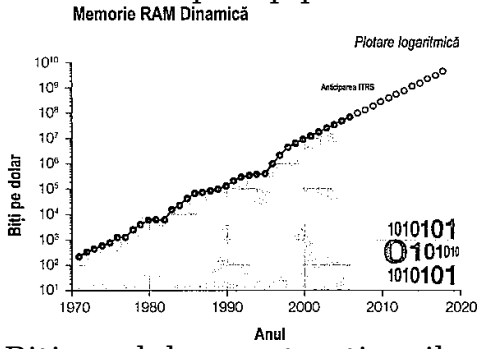
Calcule pe secundă la mia de dolari (constant) ale unor dispozitive de computare10



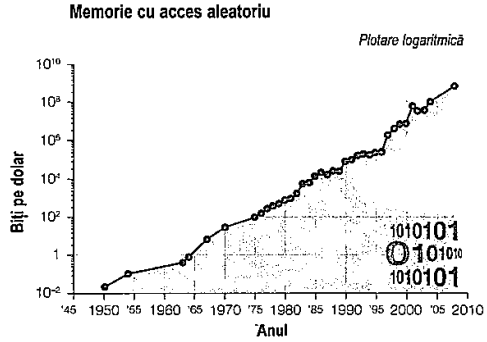
Operaţii *floating-point* pe secundă pe diverse supercomputere11



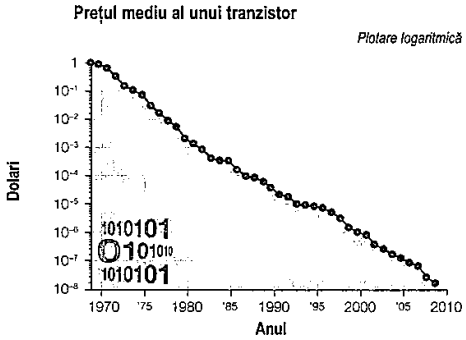
Tranzistori pe cip pentru diverse procesoare Inte112



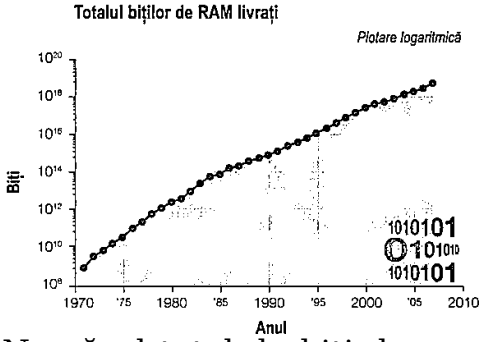
Biţi pe dolar pentru tipurile de memorie aleatorie cu acces dinamic 3



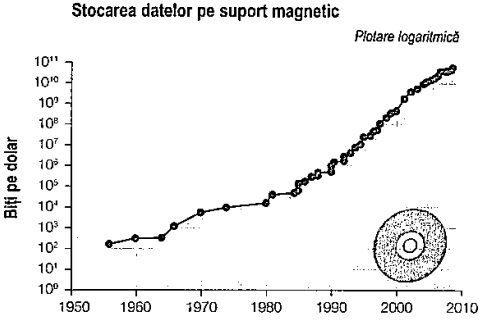
Biţi pe dolar pentru tipurile de memorie cu acces aleatoriu14



Preţul mediu al unui tranzistor, în dolari15



Numărul total de biţi de memorie cu acces aleatoriu livraţi anual16



Biţi/dolar (la valoarea constantă de 2000 USD) pentru suportul de stocare magnetică a datelor17

spirituale şi *The Singularity Is Near-Singularitatea este aproape*. Poţi citi eseul aici, mergând la link-ul de la această notă de final.9 *Epoca maşinilor spirituale* a conţinut sute de predicţii pentru următoarele decenii:

2009, 2019, 2029 şi 2099. De exemplu, în *The Age of Spiritual Machines-Epoca maşinilor spirituale*, pe care am scris-o în anii 1990, am făcut 147 de predicţii pentru anul 2009. Dintre acestea, 115 (78 la sută) sunt absolut corecte la sfârşitul anului 2009; predicţiile bazate pe mărimile capacităţii şi raportului preţ/performanţă pentru tehnologiile informaţiei au fost în mod deosebit precise.

Alte 12 predicţii (8 la sută) sunt „în esenţă corecte". Un total de 127 de predicţii (86 la sută) sunt corecte ori în esenţă corecte. (Deoarece predicţiile făcute aveau caracter specific pentru un întreg deceniu, o predicţie pentru anul 2009 a fost considerată „în esenţă corectă" dacă s-a împlinit în 2010 sau 2011.) încă 17 predicţii (12 la sută) sunt parţial corecte şi 3 (2 la sută) sunt greşite.

Nici predicţiile etichetate ca „greşite" nu au fost toate eronate. De exemplu, am considerat greşită predicţia mea că o să avem automobile autopilotate, cu toate că Google a făcut demonstraţii cu asemenea automobile în luna octombrie a anului 2010, patru camionete electrice fără şofer încheind cu succes un *drive-test* de 13.000 de kilometri din Italia până în China.18

Experţii din domeniu anticipează actualmente că aceste tehnologii vor fi disponibile în mod curent pentru public înainte de sfârşitul acestui deceniu.

Tehnologiile computaţiei şi comunicaţiilor, aflate într-o permanentă şi exponenţială dezvoltare, contribuie, toate, la proiectul de înţelegere şi recreare a sistemului creierului uman. Acest efort nu este un singur proiect organizat, ci, dimpotrivă, e rezultatul unei multitudini de proiecte diverse, inclusiv de realizare a modelelor detaliate ale componentelor creierului, mergând de la neuronii individuali la întregul neocortex, la cartografierea „conectomului" (conexiunile neuronale din creier), la simulări are unor regiuni ale creierului şi la multe altele. Toate acestea s-au amplificat exponenţial. Cea mai mare parte a dovezilor prezentate în această carte au devenit disponibile abia recent – de exemplu, studiul Wedeen din

anul 2012, discutat în Capitolul 4, care a dovedit modelul în grilă foarte ordonat şi „simplu” (ca să-i citez pe cercetători) pentru conexiunile de la nivelul neocortexului. Cercetătorii care au participat la acest studiu recunosc faptul că intuiţia lor (şi imaginile) au devenit fezabile ca rezultat al noii tehnologii a imaginilor de înaltă rezoluţie.

Tehnologiile de scanare a creierului se îmbunătăţesc sub aspectul rezoluţiei spaţiale şi temporale într-un ritm exponenţial. Diferitele tipuri de scanare a creierului care au fost cercetate merg de la metodele complet neinvazive ce pot fi folosite la om, la metodele mai invazive sau distructive folosite în cazul animalelor.

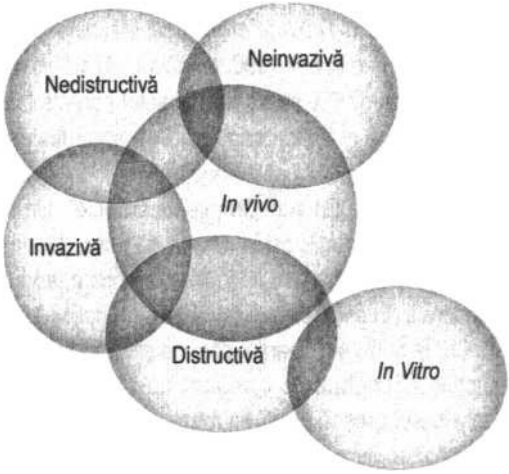
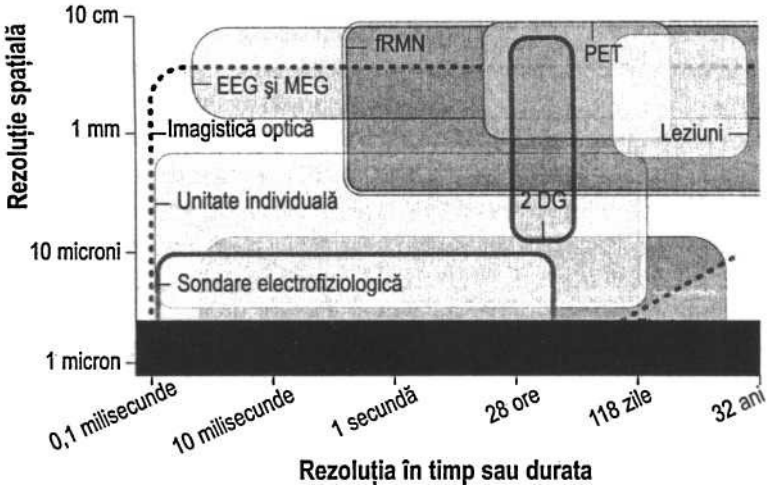
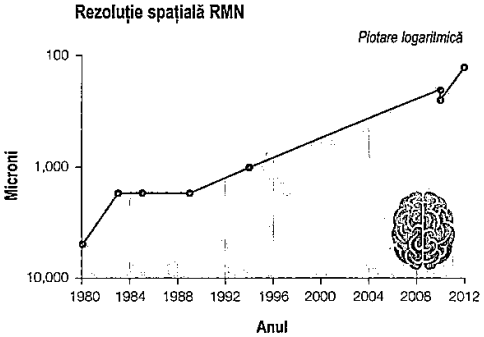


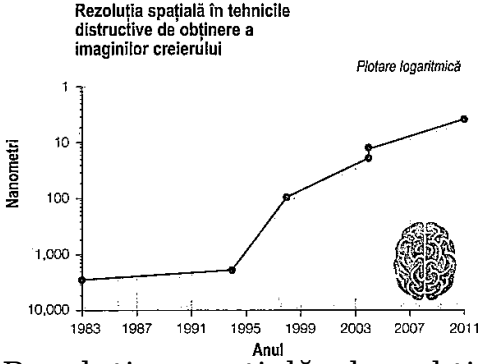
Diagrama Venn privitoare la metodele de obţinere a imaginilor creierului 19



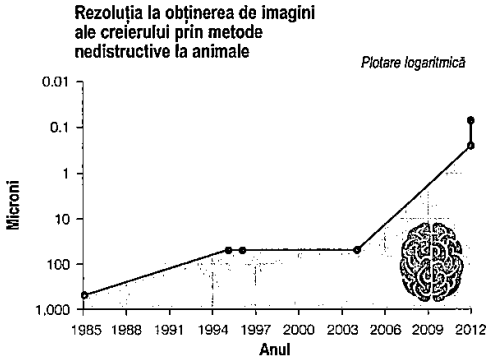
Instrumente pentru obţinerea de imagini ale creierului20



Rezoluţia spaţială RMN în microni21



Rezoluţia spaţială la obţinerea de imagini ale creierului prin metode distructive22



Rezoluţia spaţială pentru tehnicile nedistructive de obţinere a imaginilor creierului la animale23

Rezonanţa magnetică (RMN), o tehnică imagistică neinvazivă cu o rezoluţie relativ înaltă, s-a îmbunătăţit permanent, într-un ritm exponenţial, până la punctul în care rezoluţiile spaţiale sunt acum de aproape 100 de microni (a milioana parte dintr-un metru).

Tehnicile distructive de obţinere a imaginilor, care sunt aplicate pentru cartografierea conectomului (harta tuturor conexiunilor interneuronale) din creierul animalelor s-au îmbunătăţit şi ele într-un ritm exponenţial. Actuala rezoluţie maximă este în jur de patru nanometri, adică suficientă pentru a putea fi văzute conexiunile individuale.

Tehnologiile din domeniul inteligenţei artificiale, ca de pildă sistemele de înţelegere a limbii naturale, nu sunt concepute neapărat pentru imitarea principiilor teoretice ale funcţionării creierului, ci mai curând pentru o eficacitate maximă. În aceste condiţii, este notabil faptul că tehnicile care au câştig de cauză sunt consecvente principiilor pe care le-am scos în evidenţă în această carte: recognizatori autoorganizaţi, ierarhii de modele invariabile autoasociative cu redundanţă şi anticipare ascendentă & descendentă. Şi aceste sisteme se dezvoltă exponenţial, după cum a demonstrat Watson.

Un scop principal al înţelegerii creierului este extinderea arsenalului nostru de tehnici pentru crearea sistemelor inteligente. Cu toate că e posibil să nu înţeleagă

acest lucru, numeroşi cercetători din domeniul inteligenţei artificiale au fost deja influenţaţi profund de cunoştinţele pe care le deţinem cu privire la principiile de funcţionare a creierului. Înţelegerea creierului ne mai ajută şi să îndreptăm disfuncţii cerebrale de tot felul. Bineînţeles, acesta este un alt scop-cheie al proiectului de abordare a creierului prin metodele retroingineriei: să înţelegem cine suntem.

Capitolul 11 OBIECŢII

I

Dacă o maşină s-ar putea dovedi imposibil de deosebit de un om, ar trebui să-i acordăm respectul pe care l-am arăta unui om – ar trebui să acceptăm că are minte.

— Stevan Hamad

Cea mai semnificativă sursă de obiecţii la teza mea reprezentată de Legea accelerării veniturilor şi de aplicarea ei în domeniul amplificării inteligenţei umane îşi are originile în natura liniară a intuiţiei omului. După cum voi arăta în continuare, fiecare dintre cele câteva sute de milioane de recognizatori de modele de la nivelul neocortexului procesează informaţia secvenţial. Una dintre implicaţiile acestei organizări este aceea că, în privinţa viitorului, avem aşteptări liniare, prin urmare criticii îşi aplică intuiţia liniară la fenomenele informaţiei, care sunt eminamente exponenţiale.

Numesc obiecţiile de acest tip „critici din lipsă de încredere”, datorită faptului că proiecţiile exponenţiale par incredibile ţinând cont de predilecţia noastră liniară şi îmbracă o varietate de forme. Cofondatorul Microsoft Paul Allen (născut în anul 1953) şi colegul său Mark Greaves au formulat recent o serie de obiecţii de acest gen, într-un eseu intitulat *The Singularity Isn’t Ne-ar – Singularitatea nu este aproape*, publicat în revista *Technology Revlew*. Răspunsul meu de aici este destinat doar criticilor lui Allen, dar ele reprezintă o gamă tipică de obiecţii la argumentele pe care le-am avansat în special în legătură

cu creierul. În ciuda referirii lui Allen, în titlul eseului, la cartea mea *The Singularity Is Near-Singularitatea este aproape*, singurul lui citat provine dintr-un eseu pe care l-am scris în anul 2001 (intitulat *The Law of Accelerating Returns – Legea accelerării veniturilor*). Mai mult chiar, articolul lui râu recunoaşte şi nici nu ţine seamă de argumentele pe care le-am prezentat în carte. Din păcate, mult prea des constat că aşa stau lucrurile în cazul criticilor aduse cercetărilor mele.

Când a fost publicată, în anul 1999, fiind îmbogăţită ulterior cu eseul din anul 2001, cartea mea *The Age of Spiritual Machines – Epoca maşinilor spirituale* a generat critici pe mai multe direcţii, cum ar fi acelea că: Legea lui Moore va ajunge la sfârşit; capacitatea hardware se va perfecţiona exponenţial, dar partea software e împotmolită; creierul este mult prea complicat; există capacităţi ale creierului care, în mod Inerent, nu pot fi reproduse de un software; şi alte câteva. Unul dintre motivele pentru care am scris cartea *The Singularity Is Neara* fost să răspund acestor critici.

Nu pot spune că Allen şi alţi critici asemenea lui ar fi trebuit neapărat să se lase convinşi de argumentele pe care le-am prezentat în această carte, dar, cel puţin, şi el, şi ceilalţi ar fi putut răspunde la ceea ce am scris eu efectiv. Allen susţine că „Legea accelerării veniturilor… nu este o lege fizică”. Aş scoate în evidenţă faptul că partea cea mai însemnată a legilor din ştiinţă nu sunt legi fizice; ele rezultă din proprietăţile emergente ale unui mare număr de evenimente petrecute la nivel mai redus. Un exemplu clasic este reprezentat de Legile termodinamicii. Dacă te uiţi pe suportul matematic aflat la baza Legilor termodinamicii, observi că ele realizează pentru fiecare particulă un model care urmează o cale aleatorie, astfel încât, prin definiţie, nu putem anticipa unde anume se va afla o anumită particulă într-un moment din viitor. Cu toate acestea, proprietăţile de ansamblu ale gazului sunt cât se poate de predictibile, cu un grad înalt de precizie, potrivit

Legilor termodinamicii. La fel se întâmplă şi cu Legea accelerării veniturilor: orice proiect şi contribuitor din domeniul tehnologic e imprevizibil, însă traiectoria de ansamblu, aşa cum este ea cuantificată de mărimea raportului preţ/performanţă şi capacitate, urmează, totuşi, o cale remarcabil de predictibilă.

Dacă tehnologia computerelor ar reprezenta preocuparea unui grup restrâns de cercetători, ea ar fi într-adevăr imprevizibilă. Însă este produsul unui sistem de proiecte suficient de dinamic pentru ca raportul fundamental preţ/performanţă (de exemplu calcule pe secundă per dolar constant) să urmeze o cale exponenţială foarte netedă, datând încă de la recensământul american din anul 1890, aşa cum am subliniat în capitolul anterior. Deşi baza teoretică pentru Legea accelerării veniturilor este prezentată pe larg în cartea mea *The Singularity Is Near-Singularitatea este aproape*, demonstraţiile cele mai solide în sprijinul ei vin din dovezile empirice pe care le-am prezentat eu şi alţii.

Allen scrie că „aceste «legi» merg până nu mai merg”. Aici, el confundă paradigmele cu traiectoria continuă a unui domeniu de bază al tehnologiei informaţiei. Dacă am examina, de exemplu, tendinţa construirii unor tuburi cu vid din ce în ce mai mici – paradigma progresului computaţiei în anii 1950 am spune că e adevărat că ea a mers până nu a mai mers. Dar, atunci când finalul acestei paradigme a devenit evident, a crescut presiunea la care a fost supusă cercetarea pentru emiterea următoarei paradigme. Tehnologia tranzistorilor a păstrat tendinţa de creştere exponenţială a raportului preţ/performanţă în domeniul computaţiei şi asta a condus la a cincea paradigmă (Legea lui Moore) şi la continua comprimare a particularităţilor pe circuite integrate. S-au emis permanent predicţii referitoare la faptul că Legea lui Moore va ajunge la un final. Setul de documente intitulat *International Technology Roadmap for Semiconductors – Ghidul tehnologic International pentru semiconductori*

proiectează particularităţi de şapte nanometri până la începutul anilor ’2020.2 în acest punct, particularităţile-cheie vor avea diametrul de treizeci şi cinci de atomi de carbon şi va fi dificil să fie comprimate şi mai mult. Pe de altă parte însă, Intel şi alţi producători de cipuri întreprind deja primii paşi în direcţia celei de-a şasea paradigme, computarea în trei dimensiuni, pentru continuarea îmbunătăţirii exponenţiale a raportului preţ/performanţă. Intel intenţionează să transforme cipurile tridimensionale în ceva obişnuit într-un orizont de timp de 13 – 19 ani; deja au fost lansaţi tranzistor» tridimensionali şi cipurile de memorie 3D. Această a şasea paradigmă va continua să crediteze Legea accelerării veniturilor în privinţa raportului preţ/performanţă până mai târziu în decursul acestui secol, când un volum de computaţie în valoare de o mie de dolari va fi de miliarde de ori mai puternic decât creierul omului.3 (Se pare că Allen şi cu mine ne punem în sfârşit de acord în privinţa nivelului de computaţie necesar pentru simularea funcţională a creierului uman.) 4

Allen merge apoi mai departe şi avansează argumentul-standard că software-ul nu progresează în aceeaşi manieră exponenţială ca hardware-ul. În cartea mea *The Singularity Is Near-Singularitatea este aproape*, m-am ocupat pe larg de acest subiect, citând diverse metode de măsurare a complexităţii şi capacităţii programelor software, care într-adevăr demonstrează o creştere exponenţială similară.5 Un studiu recent (*Report to the President and Congress*, *Designing a Digital Future*: *Federally Funded Research and Development în tjetworking and Information Technology – Raport pentru preşedinte şi Congres privind conceperea unui viitor digital: cercetare şi dezvoltare finanţată de la nivel federal pentru networking şi tehnologia informaţiei* – aparţinând grupului de consilieri ai preşedintelui pentru ştiinţă şi tehnologie) afirmă următoarele:

Şi mai remarcabil – şi chiar mai puţin înţeles – este faptul că, în numeroase domenii, performanţa are câştig de

cauză pentru că perfecţionarea algoritmilor a depăşit cu mult chiar şi performanţele spectaculoase datorate creşterii vitezei procesoarelor. Algoritmii pe care îi folosim astăzi pentru recunoaşterea vorbirii, pentru traducerea în limbă naturală, pentru jocul de şah sau pentru planificarea logistică au evoluat remarcabil în ultimul deceniu… Iată un singur exemplu, oferit de profesorul Martin Grotschel de la Konrad-Zuse-Zentrum fur Informations technik din Berlin. Grotschel, expert în optimizare, observă că realizarea unui model de planificare pentru o linie de producţie cu ajutorul programării liniare ar fi avut nevoie de 82 de ani în 1988, folosindu-se computerele şi algoritmii de programare liniară ai vremii. Cincisprezece ani mai târziu – în anul 2003 –, acelaşi model putea fi realizat în doar circa un minut, ceea ce înseamnă o îmbunătăţire de aproximativ 43 de milioane de ori. Din această sumă, o creştere de aproximativ 1.000 de ori s-a datorat măririi vitezei procesorului, iar restul de 43.000 de ori a fost datorată progreselor privitoare la algoritmi! Grotschel mai citează şi o îmbunătăţire de circa 30.000 de ori a algoritmilor pentru programarea mixtă a întregilor între anii 1991 şi 2008. Proiectarea şi analiza algoritmilor, ca şi studiul complexităţii computaţionale inerente a problemelor sunt subdomenii fundamentale ale ştiinţei computerelor.

Reţine, cititorule, că programarea liniară, citată mai înainte de Grotschel ca fiind beneficiara unei îmbunătăţiri a performanţelor de 43 de milioane la 1, este tehnica matematică folosită pentru atribuirea optimă a resurselor într-un sistem de memorie ierarhică de felul Modelelor Markov Ascunse Ierarhic despre care am vorbit anterior. Şi citez şi alte asemenea exemple în cartea mea *The Singularity Is Ne-ar.6*

În privinţa inteligenţei artificiale, Allen se grăbeşte să-l respingă pe Watson, creaţia IBM – opinie împărtăşită de numeroşi alţi critici. Mulţi dintre aceşti detractori nu ştiu absolut nimic despre Watson în afara faptului că e un software care rulează pe un computer (deşi este unul

paralel cu 720 de procesoare). Allen scrie că sisteme precum Watson „rămân fragile, limitele performanţelor lor sunt stabilite rigid de premisele şi algoritmii lor definitorii interni şi adeseori dau răspunsuri fără sens în afara domeniilor lor specifice”.

În primul rând, am putea face o observaţie similară în privinţa omului. De asemenea, aş sublinia că „domeniile specifice” ale lui Watson includ toată Wikipedia, plus multe alte baze de cunoştinţe care cu greu ar putea fi considerate un obiectiv îngust. Watson se ocupă de o gamă bogată de cunoştinţe umane şi este capabil să facă faţă unor forme subtile ale limbajului, inclusiv jocuri de cuvinte, comparaţii şi metafore, practic pentru orice domeniu al strădaniilor omeneşti. Nu este perfect, darnici oamenii nu sunt, iar el a fost suficient de bun ca să câştige la Jeopardy Hn faţa celor mai buni jucători umani.

Allen susţine că Watson a fost asamblat personal de către oamenii de ştiinţă, care au construit fiece legătură de cunoaştere îngustă în domenii specifice. Pur şi simplu, acest fapt nu este adevărat. Deşi câteva domenii ale datelor lui Watson au fost programate direct, Watson şi-a dobândit singur majoritatea semnificativă a cunoştinţelor citind în limbaj natural documente cum ar fi Wikipedia. Acesta este punctul-forte al lui Watson, iar lui i se adaugă capacitatea de a înţelege limbajul ocolit al întrebărilor de la *Jeopardy*! (răspunsuri în căutarea unei întrebări).

După cum am menţionat anterior, în mare, criticile la adresa lui Watson vizează faptul că acesta funcţionează pe baza probabilităţilor statistice şi nu a unei „adevărate” înţelegeri. Mulţi cititori pricep că asta înseamnă că Watson nu face decât să adune statistici privitoare la succesiunile cuvintelor. Sintagma „informaţie statistică”, în cazul lui Watson, se referă de fapt la coeficienţii distribuiţi şi la conexiunile simbolice ale metodelor autoorganizate, cum ar fi Modelele Markov Ascunse Ierarhice. La fel de uşor se poate respinge ideea potrivit căreia concentraţiile neurotransmiţătorilor şi modelele de conexiuni redundante

de la nivelul cortexului uman sunt „informaţii statistice”. Da, şi noi rezolvăm ambiguităţile într-un mod foarte asemănător cu Watson – apreciind probabilitatea diverselor interpretări ale unei expresii.

Allen continuă: „Orice structură [din creier] a fost precis adecvată de milioanele de ani de evoluţie pentru a face un anumit lucru, indiferent care este. acesta. Nu e ca un computer, cu milioane de tranzistori identici în dispoziţii regulate de memorie controlate de o unitate centrală de procesare cu câteva elemente diferite. La nivelul creierului, toate structurile individuale şi toate circuitele neuronale au fost rafinate individual de evoluţie şi de factorii de mediu”.

Această afirmaţie că toate structurile şi toate circuitele neuronale ale creierului sunt unice şi se află intenţionat acolo este pur şi simplu imposibilă, căci ar însemna că proiectul creierului ar impune sute de mii de miliarde de biţi de informaţie. Planul structural al creierului (asemenea celui al restului corpului) este conţinut în genom şi creierul însuşi nu poate să depoziteze mai multă informaţie de concepţie decât genomul. Observă, cititorule, că informaţia epigenetică (de pildă controlul expresiei genelor de către peptide) nu contribuie în mod apreciabil la volumul de informaţie de la nivelul genomului. Experienţa şi învăţarea sporesc într-adevăr semnificativ volumul de informaţie conţinut de creier, dar se pot spune şi despre inteligenţa artificială aceleaşi lucruri care se spun despre Watson. În cartea mea *The Singularity is Near-Singularitatea este aproape*, arăt că, după comprimarea fără pierderi (datorată redundanţei masive a genomului), volumul de informaţie de concepţie din genom se ridică la aproximativ 50 de milioane de biţi, din care în jur de jumătate (respectiv circa 25 de milioane de biţi) are legătură cu creierul.7 Nu este ceva simplu, însă e un nivel de complexitate căruia îi putem face faţă şi reprezintă o complexitate mai redusă decât a multor sisteme software din lumea modernă. În plus, peste 25 de

milioane de biţi din informaţia genetică de concepţie a creierului au legătură cu cerinţele biologice pentru neuroni, nu cu algoritmii lor de funcţionare.

Cum ajungem la un număr de 100 până la 1.000 de mii de miliarde de conexiuni la nivelul creierului de pe urma câtorva zeci de milioane de biţi de informaţie de concepţie? în mod evident, răspunsul este: prin redundanţă masivă. Dharmendra Modha, managerul de Computaţie Cognitivă de la IBM Research, scrie că „neuroanatomiştii nu au descoperit *o* reţea incredibil de încâlcită, arbitrar conectată, cu totul deosebită, în creierul fiecărui individ, ci, dimpotrivă, foarte multe structuri care se repetă şi o mare analogie între specii… Reconfigurabilitatea naturală uluitoare dă speranţe că algoritmii de bază pentru neurocomputaţie sunt independenţi de modalităţile senzoriale şi motorii şi că mare parte din variaţiile observate la nivelul structurii corticale a tuturor regiunilor reprezintă o rafinare a unui circuit canonic; şi tocmai acest circuit canonic este cel pe care vrem să-l copiem prin retroinginerie".8

Allen pledează în favoarea unei inerente „frâne de complexitate, care în mod necesar ar limita progresul înţelegerii creierului uman şi replicarea capacităţilor lui", pe baza ideii că toate cele aproximativ 100 până la 1.000 de mii de miliarde de conexiuni din creierul uman există datorită unui proiect explicit. Expresia sa „frână de complexitate" confundă pădurea cu copacii. Dacă vrei să înţelegi, să realizezi modele, să simulezi şi să recreezi un pancreas, nu trebuie să refaci sau să simulezi fiecare organită din fiecare celulă a insulelor pancreatice Langerhans. Ai vrea, în schimb, să înţelegi o asemenea celulă, să-i abstractizezi funcţionalitatea fundamentală legată de controlul nivelului insulinei şi pe urmă să o extinzi la un grup numeros de asemenea celule. Acest algoritm este bine înţeles în ceea ce priveşte celulele din insulele Langerhans. Actualmente există pancreasuri artificiale ce utilizează acest model funcţional care este

testat. Deşi cu siguranţă există o complexitate şi o variabilitate cu mult mai mari la nivelul creierului comparativ cu complexitatea şi variabilitatea celulelor din insulele Langerhans ale pancreasului, se constată pe de altă parte şi o masivă repetare a funcţiilor, aşa cum am arătat adesea în această carte.

Criticii care merg pe linia lui Allen mai manifestă şi ceea ce eu aş numi „pesimismul omului de ştiinţă”. Cercetătorii implicaţi în lucrul la următoarea generaţie a unei tehnologii sau a modelelor dintr-un domeniu ştiinţific se luptă în mod invariabil cu acel set imediat de provocări, astfel încât, dacă cineva descrie felul cum va arăta tehnologia peste zece generaţii, le sclipesc ochii. Unul dintre pionierii circuitelor integrate îmi reamintea recent de strădaniile de trecere de la mărimi ale particularităţilor de 10 microni (10.000 de nanometri) la mărimi de 5 microni (5.000 de nanometri), în urmă cu mai bine de treizeci de ani. Oamenii de ştiinţă erau încrezători cu prudenţă în atingerea acestui scop, dar, când au anticipat că într-o bună zi vom avea într-adevăr circuite cu dimensiunile particularităţilor sub un micron (1.000 de nanometri), cei mai mulţi dintre ei, concentraţi asupra scopului, s-au gândit că era un lucru prea exagerat ca să se oprească asupra lui. Au fost ridicate obiecţii referitoare la fragilitatea circuitelor la acel nivel de precizie, la efectele termice şi aşa mai departe. Astăzi, Intel începe să folosească cipuri cu lungimi de poartă de 22 de nanometri.

Am fost martorii aceluiaşi gen de pesimism în privinţa Proiectului Genomului Uman. La jumătatea efortului cu durata de cincisprezece ani, doar 1 la sută din genom fusese achiziţionat, iar criticii propuneau limite elementare privind viteza cu care putea fi secvenţializat acesta fără distrugerea delicatelor structuri genetice. Însă, mulţumită creşterii exponenţiale a capacităţii şi a raportului preţ/performanţă, proiectul a fost încheiat şapte ani mai târziu. Proiectul de copiere a creierului uman prin retroinginerie progresează în mod similar. Abia recent, de

exemplu, am ajuns cu tehnicile neinvazive la un prag care să ne permită să observăm în timp real conexiunile interneuronale individuale atunci când se formează şi generează impulsurile nervoase. Cele mai multe dintre dovezile pe care le-am prezentat în cartea de faţă au depins de aceste evoluţii şi abia recent au devenit disponibile.

Alien mi-a descris propunerea de recreare a creierului uman prin retroinginerie ca pe o simplă scanare a creierului pentru înţelegerea structurii sale fine, urmată de simularea întregului creier „pe invers”, fără înţelegerea metodelor lui de procesare a informaţiilor. Trebuie neapărat să înţelegem în detaliu cum funcţionează diferitele tipuri individuale de neuroni şi abia apoi să adunăm informaţia cu privire la conectarea modulelor funcţionale. Metodele derivate din acest gen de analiză pot apoi să ghideze dezvoltarea unor sisteme inteligente. În esenţă, căutăm metode inspirate de biologie care pot accelera munca în domeniul inteligenţei artificiale. Această muncă a progresat în mare parte în absenţa unor intuiţii semnificative privitoare la modul în care creierul îndeplineşte funcţii similare. Din cercetările mele personale în domeniul recunoaşterii vorbirii, ştiu că munca ne-a fost accelerată foarte mult din momentul când am beneficiat de o intuiţie referitoare la modul în care creierul pregăteşte şi transformă informaţia auditivă.

Modul în care se diferenţiază structurile masiv redundante din creier rezidă în învăţare şi experienţă. Vârful actual al inteligenţei artificiale le permite, de fapt, sistemelor să înveţe şi din propria experienţă. Automobilele Google care se pilotează singure învaţă din propria lor experienţă de pilotaj şi din datele automobilelor Google pilotate de şoferi umani; Watson şi-a acumulat cea mai mare parte a cunoştinţelor citind singur. Este interesant de observat că metodele folosite astăzi în domeniul inteligenţei artificiale au evoluat pentru a fi matematic foarte similare cu mecanismele de la nivelul

neocortexului.

O altă obiecţie la adresa fezabilităţii „inteligenţei artificiale puternice” (inteligenţă artificială la nivel uman şi peste) susţine că creierul uman foloseşte extensiv computares analogică, în vreme ce metodele digitale nu pot replica în mod inerent gradaţiile de valoare pe care le pot întruchipa reprezentările analogice. Este adevărat că un bit este fie deschis, fie închis, însă cuvintele compuse din mai mulţi biţi reprezintă cu uşurinţă mai multe gradaţii şi o pot face cu orice grad de acurateţe dorit. Desigur, computerele digitale fac acest lucru permanent. Ca urmare, acurateţea informaţiei analogice la nivelul creierului (tăria sinaptică, de exemplu) este doar de un nivel dintre cele 256 de niveluri de opt biţi care pot fi reprezentate.

În Capitolul 9, am citat obiecţia lui Roger Penrose şi Stuart Hameroff referitoare la microtubuli şi la computares cuantică. Îţi aminteşti, cititorule: ei afirmau că structurile de microtubuli din neuroni realizează o computare cuantică, aşa încât, de vreme ce nu este posibil să faci asta în cazul computerelor, creierul uman e fundamental diferit şi, după câte s-ar putea presupune, mai bun. Aşa cum am argumentat mai înainte, nu există nicio dovadă că microtubulii neuronali realizează computarea cuantică. De fapt, omul se descurcă destul de slab la rezolvarea genului de probleme la care un computer cuantic ar excela (cum ar fi înmulţirea unor numere mari). Şi, dacă s-ar dovedi a fi adevărat ceva din toate acestea, nimic nu ar putea să împiedice folosirea computării cuantice şi în cazul computerelor.

John Searle este celebru pentru lansarea experimentului mintal pe care el îl numeşte „camera chinezească”, un argument pe care îl discut detaliat în cartea mea *The Singularity is Ne-ar.9* Pe scurt, este vorba despre un bărbat care primeşte întrebări scrise în limba chineză şi apoi răspunde la ele. Pentru asta, foloseşte o carte complicată de reguli. Searle susţine că bărbatul nu

înţelege cu adevărat limba chineză şi că nu este „conştient” în limba respectivă (deoarece nu înţelege întrebările şi răspunsurile), în ciuda aparentei sale abilităţi de a răspunde la întrebări. Searle compară această situaţie cu cea a unui un computer şi concluzionează că un computer care ar putea răspunde la întrebări puse în chineză (în esenţă trecând un test Turing în limba chineză), asemenea bărbatului din camera chinezească, nu ar înţelege în mod real limba şi nici nu ar fi conştient de ceea ce face.

Există mai multe scăpări filosofice în argumentaţia lui Searle. În primul rând, bărbatul din acest experiment mintal este comparabil doar cu unitatea centrală de procesare a unui computer. S-ar putea spune că unitatea centrală de procesare nu înţelege cu adevărat ce face, însă ea e doar o parte a structurii. În camera chinezească a lui Searle, bărbatul cu cartea de reguli în mână constituie întregul sistem. Acest sistem înţelege limba chineză; altminteri, nu ar fi capabil să răspundă convingător la întrebările puse în această limbă, fapt ce ar viola ipoteza lui Searle din experimentul său mintal.

Atractivitatea argumentului lui Searle rezidă în faptul că, în zilele noastre, este dificil să se deducă ideea că un software ar înţelege cu adevărat sau ar avea conştienţă. Însă problema acestui argument este că poţi să aplici raţionamentul său şi creierului uman. Toţi recognizatorii neocorticali de modele – şi chiar toţi neuronii şi toate componentele fiecărui neuron – respectă un algoritm. (La urma urmelor, sunt mecanisme moleculare, care respectă legea naturală.) Dacă am concluziona că respectarea unui algoritm este incompatibilă cu adevărata înţelegere şi conştienţă, atunci ar trebui să concluzionăm că nici creierul uman nu prezintă aceste calităţi. Poţi lua argumentul lui Searle cu camera chinezească şi pur şi simplu poţi înlocui sintagma lui „manipularea simbolurilor” cu „manipularea conexiunilor interneuronale şi a tăriei sinaptice”. Vei obţine un argument cu adevărat

convingător în sensul că creierul uman realmente nu poate să înţeleagă nimic.

O altă direcţie a argumentării vine din însăşi natura naturii, care a devenit un nou teren sacru pentru mulţi observatori. De exemplu, biologul neo-zeelandez Michael Denton (născut în anul 1943) vede o diferenţă profundă între principiile de concepere a maşinilor şi cele ale biologiei. Denton scrie că entităţile naturale sunt „autoorganizate… autoreferenţiale… autoreplicative… autoformatoare şi… holistice".10 El susţine că asemenea forme biologice nu pot fi create decât prin procese biologice şi că, prin urmare, ele sunt realităţi „imuabile… impenetrabile şi… fundamentale" ale existenţei, ceea ce înseamnă că reprezintă o categorie filosofică diferită de a maşinilor.

Realitatea, aşa cum am văzut, este că pot fi concepute şi maşini pe baza aceloraşi principii. A învăţa paradigmele specifice de concepţie ale celei mai inteligente entităţi a naturii – creierul uman – este exact scopul proiectului de realizare a creierului prin retroinginerie. De asemenea, e neadevărat faptul că sistemele biologice sunt complet „holistice", după cum se exprimă Denton şi nici că, reciproc, maşinile trebuie să fie complet modulare. Am identificat în mod clar ierarhii ale unităţilor de funcţionalitate în sistemele naturale, în special în creier, iar sistemele de inteligenţă artificială folosesc metode comparabile.

Îmi face impresia că mulţi critici vor rămâne nemulţumiţi până când computerele vor trece Testul Turing în mod curent, însă nici acest prag nu va fi clar. Fără îndoială, vor apărea controverse legate de faptul că pretinsele Teste Turing administrate ar fi valide. Fără îndoială, mă voi număra probabil printre acei critici care vor discredita afirmaţiile timpurii pe aceste direcţii. Până se vor linişti disputele referitoare la validitatea unui computer care trece Testul Turing, computerele vor fi depăşit de mult inteligenţa umană neasistată.

Accentul meu aici este pus pe cuvântul „neasistată”, deoarece asistarea este exact motivul pentru care creăm aceşti „copii ai minţii”, cum îi numeşte Hans Moravec.11 Combinarea recogniţiei de modele de nivel uman cu viteza inerentă şi acurateţea computerelor se va concretiza prin abilităţi redutabile. Însă nu este vorba despre o invazie extraterestră a maşinilor inteligente de pe Marte – creăm aceste instrumente pentru a deveni noi înşine mai inteligenţi. Cred că majoritatea observatorilor vor fi de acord cu mine că acesta este aspectul unic al speciei umane: construim aceste instrumente pentru a ne extinde propria anvergură.

EPILOGUL

Imaginea e foarte anostă, domnilor… în lume clima e în schimbare, mamiferele preiau supremaţia şi avem cu toţii un creier cât o nucă.

— *Dinozaurii care vorbesc*, în *The Far Side* de Gary Larson

Inteligenţa poate fi definită drept abilitatea de a rezolva probleme cu ajutorul unor resurse limitate, dintre care o resursă-cheie o reprezintă timpul. Prin urmare, abilitatea de a rezolva mai rapid o problemă cum ar fi găsirea hranei sau evitarea unui prădător reflectă o forţă mai mare a intelectului. Inteligenţa a evoluat deoarece a fost utilă pentru supravieţuire – fapt ce poate părea evident, dar cu care nimeni nu este de acord. Aşa cum a fost aplicată de specia noastră, inteligenţa ne-a permis nu numai să dominăm planeta, ci şi să ne îmbunătăţim permanent calitatea vieţii. Nici acest din urmă aspect nu este evident pentru toată lumea, dat fiind faptul că astăzi există percepţia larg răspândiră că viaţa devine din ce în ce mai grea. De exemplu, un sondaj Gallup dat publicităţii în data de 4 mai 2011 a dezvăluit că numai „44 la sută dintre americani cred că tinerii de astăzi vor avea o viaţă mai bună decât părinţii lor”.1

Dacă ne uităm la tendinţele mai ample, nu numai că speranţa de viaţă a omului s-a mărit de patru ori pe

parcursul ultimului mileniu (şi a crescut de mai mult de două ori în cursul ultimelor două secole) 2, dar şi PIB-ul pe cap de locuitor (în dolari constanţi la valoarea curentă) a crescut de la câteva sute de dolari în anul 1800, la mii de dolari în zilele noastre, cu tendinţe şi mai pronunţate în lumea dezvoltată.3 în urmă cu un secol, existau doar câteva democraţii, în vreme ce azi ele reprezintă standardul. Pentru a căpăta o perspectivă istorică privitoare la proporţiile progresului nostru, îţi sugerez să citeşti *Leviatanul* lui Hobbes (1651), în care autorul descrie „viaţa omului" ca fiind „solitară, săracă, grea, brutală şi scurtă". Pentru o perspectivă modernă, recenta carte *Abundance* (2012) a întemeietorului X-Prize Foundation (şi cofondator alături de mine al Singularity University) Peter Diamandis şi a autorului pe teme ştiinţifice Steven Kotler documentează modurile extraordinare în care viaţa din zilele noastre s-a îmbunătăţit permanent pe toate dimensiunile ei. Recenta carte a lui Steven Pinker *The Better Angels of Our Nature*: *Why Violence Has Declined* (2011) probează minuţios amplificarea relaţiilor paşnice dintre oameni şi popoare. Avocatul, antreprenorul şi autorul american Martine Rothblatt (născută în anul 1954) documentează îmbunătăţirea permanentă a drepturilor civile, subliniind, de exemplu, că, în câteva decenii, căsătoriile homosexuale au ajuns de la stadiul de a nu fi acceptate niciunde în lume la recunoaşterea legală într-un număr tot mai mare de jurisdicţii.4

Un motiv principal pentru care oamenii cred că viaţa din zilele noastre este din ce în ce mai grea rezidă în faptul că informaţiile privitoare la problemele lumii s-au amplificat permanent. În zilele noastre, dacă are loc o bătălie undeva pe Pământ, trăim experienţa ei aproape ca şi cum ne-am afla acolo. În timpul celui de-al Doilea Război Mondial, puteau să piară zeci de mii de oameni într-o bătălie şi, în cazul în care publicul ajungea să afle, acest lucru se întâmpla odată cu vizionarea unui jurnal plin de

purici, într-un cinematograf, ia distanţă de câteva săptămâni. În timpul Primului Război Mondial, o elită redusă numeric putea să citească despre evoluţia conflictului în ziar (fără imagini). În secolul al nouăsprezecelea, aproape că nu exista acces la ştiri de actualitate.

Progresul pe care l-am înregistrat ca specie datorită inteligenţei noastre este reflectat în evoluţia cunoaşterii, care include tehnologia şi cultura noastră. Diversele tehnologii ale omului devin într-o măsură din ce în ce mai mare tehnologii ale informaţiei, continuând inerent progresul într-o manieră exponenţială. Prin asemenea tehnologii, putem răspunde marilor provocări ale omenirii, cum ar fi păstrarea unui mediu sănătos, asigurarea resurselor pentru creşterea populaţiei (inclusiv energie, alimente şi apă), prevenirea şi vindecarea bolilor, creşterea amplă a longevităţii omului şi eradicarea sărăciei. Numai extinzându-ne cu tehnologii inteligente putem atinge complexitatea necesară pentru înfruntarea acestor provocări.

Aceste tehnologii nu sunt avangarda unei invazii inteligente care va concura cu noi şi în cele din urmă ne va disloca. De când am pus mâna pe un băţ ca să ajungem la o creangă mai de sus, ne-am folosit uneltele pentru a ne extinde anvergura, atât fizic, cât şi mental. Faptul că astăzi putem scoate din buzunar un dispozitiv şi putem accesa cunoştinţele umane cu câteva tastări ne duce dincolo de tot ce era imaginabil pentru majoritatea observatorilor cu numai câteva decenii în urmă. „Telefonul celular” (termenul este pus în ghilimele pentru că e vorba de ceva cu mult mai mult decât de un simplu telefon) din buzunarul meu este de un milion de orí mai ieftin, dar de o mie de orí mai puternic decât computerul pe care îl împărţeau toţi profesorii şi studenţii de la MIT pe când studiam acolo. Asta înseamnă o creştere de un miliard de ori a raportului preţ/performanţă în decursul ultimilor patruzeci de ani, ascensiune ia care vom fi din nou martori în următorii

douăzeci şi cinci de ani, când ceea ce încăpea într-o clădire şi acum ne încape în buzunar va încăpea într-un element figurat sanguin.

În acest fel, vom fuziona cu tehnologia inteligentă pe care o creăm. Nanoboţii inteligenţi din circulaţia noastră sanguină ne vor menţine corpurile biologice sănătoase la nivel celular şi molecular. Vor pătrunde în creier în mod neinvaziv, prin capilare şi vor interacţiona cu neuronii noştri biologici, amplificându-ne în mod direct inteligenţa. Nu este ceva atât de futurist pe cât ar putea părea. Există numeroase dispozitive de mărimea unui element figurat al sângelui care pot să vindece diabetul de tipul I la animale sau să detecteze şi să distrugă celulele canceroase din circulaţia sanguină. Pe baza Legii accelerării veniturilor, aceste tehnologii vor fi de un miliard de ori mai eficiente în următoarele trei decenii decât sunt acum.

Realmente consider dispozitivele pe care le folosesc şi norul de resurse de computare la care sunt ele conectate virtual ca fiind extensii ale mele şi mă simt incomplet dacă mă văd despărţit de aceste prelungiri ale creierului, latămotivul pentru care greva de o zi de la Google, Wikipedia şi de la mii de alte website-uri împotriva SOPA (Stop Online Piracy Act – Stop Legii Pirateriei Online) din data de 18 ianuarie 2012 a fost atât de însemnată: am avut sentimentul că b parte din creierul meu se afla în grevă (cu toate că eu şi alţii am găsit totuşi modalităţi de a accesa resursele online). A mai fost şi o demonstraţie impresionantă a forţei politice a acestor site-uri: s-a renunţat imediat ia respectiva lege – care părea că se află în linie dreaptă pentru ratificare. Dar, fapt mai important, evenimentul a arătat cât de amplu ne-am externalizat deja părţi din propria gândire în norul de computare. Deja, el face parte din ceea ce suntem. Din momentul în care vom deţine în mod curent inteligenţă nonbiologică în creierul nostru, această amplificare – şi norul la care e conectată ea – va continua să crească exponenţial din punctul de vedere al capacităţii.

Inteligenţa pe care o vom crea prin intermediul copierii creierului cu ajutorul retroingineriei va avea acces la propriul cod-sursă şi va putea să se îmbunătăţească pe sine rapid, într-un ciclu recurent de proiectare în continuă accelerare. Deşi creierul biologic uman beneficiază de o plasticitate considerabilă, după cum am văzut, el are şi o arhitectură relativ fixă, care nu poate fi modificată în mod semnificativ, plus o capacitate limitată. Nu suntem capabili să creştem numărul de 300 de milioane de recognizatori de modele la 400 de milioane, să spunem, dacă nu o facem nonbiologic. Din momentul în care vom putea realiza aşa ceva, nu va mai exista niciun motiv pentru care să ne oprim la un anumit nivel de capacitate. Putem continuă până la a face un miliard de recognizatori de modele sau o mie de miliarde.

Din perfecţionarea cantitativă rezultă progresul calitativ. Cel mai important avans al evoluţiei în cazul lui *Homo sapiens* a fost cantitativ: dezvoltarea unei frunţi mai mari, care să adăpostească un neocortex mai mare. Capacitatea sporită a neocortexului a permis acestei noi specii să creeze şi să cultive gânduri la niveluri conceptuale superioare, fapt concretizat în consacrarea tuturor variatelor domenii ale artei şi ştiinţei. Cu cât adăugăm mai mult neocortex la o formă nonbiologică, cu atât ne putem aştepta la niveluri calitative şi mai înalte ale abstracţiei.

Matematicianul britanic Irvin J. Good, coleg cu Alan Turing, a scris în anul 1965 că „prima maşină ultrainteligentă este ultima invenţie pe care trebuie s-o mai facă omul”. El a definit o asemenea maşină ca fiind una ce poate să depăşească „activităţile intelectuale ale oricărui om, indiferent dă cât de inteligent e acesta” şi a concluzionat că „de vreme ce crearea maşinilor este una dintre aceste activităţi intelectuale, o maşină inteligentă ar putea să proiecteze maşini şi mai inteligente; pe urmă, se va produce fără îndoială o «explozie de inteligenţă»”.

Ultima invenţie pe care a trebuit s-o facă evoluţia

biologică – neocortexul – conduce inevitabil la ultima invenţie pe care trebuie s-o facă omenirea – maşinile cu adevărat inteligente –, iar proiectul uneia inspiră proiectul celeilalte. Evoluţia biologică mai continuă, însă evoluţia tehnologică se produce de un milion de ori mai rapid decât aceasta. Potrivit Legii accelerării veniturilor, până la sfârşitul acestui secol, vom putea să creăm computaţie la limitele posibilului, pe baza legilor fizicii, aşa cum sunt ele aplicate computaţiei.5 Materia şi energia organizate astfel sunt numite „computronium" şi formează o combinaţie mult mai puternică sub toate aspectele decât creierul uman. Nu va fi vorba doar de computare brută, aceasta urmând să fie infuzată cu algoritmii inteligenţi care constituie întreaga cunoaştere om-maşină. În timp, vom converti mare parte din masa şi energia micului nostru colţ de galaxie, care este adecvat pentru acest scop, în computronium. Pe urmă, pentru a (¡ne în stare de funcţiune Legea accelerării veniturilor, va trebui să ne răspândim în restul galaxiei şi al Universului.

Dacă viteza luminii rămâne într-adevăr o limită inexorabilă, colonizarea Universului va dura foarte mult timp, dat fiind faptul că sistemul solar cel mai apropiat de Pământ se află la o distanţă de patru ani-lumină de acesta. Dacă există fie şi cele mai subtile mijloace de ocolire a acestei limite, inteligenţa şi tehnologia noastră vor fi suficient de puternice pentru a le exploata. Acesta este unul dintre motivele pentru care recenta sugestie că miuonii care au străbătut 730 de kilometri de la acceleratorul CERN aflat la frontiera francoelveţiană până în laboratorul Gran Sasso din centrul Italiei şi au părut să se deplaseze mai rapid decât viteza luminii au constituit o ştire potenţial importantă. Această observaţie pare să fie o falsă alarmă, dar există şi alte posibilităţi de ocolire a acestei limite. Nici măcar nu trebuie să depăşim viteza luminii dacă putem identifica scurtături către locuri aparent depărtate prin alte dimensiuni ale spaţiului în afara celor trei cu care suntem familiarizaţi. Dacă ne

dovedim sau nu capabili să depăşim ori să ocolim în vreun fel viteza luminii ca limită, asta va fi o chestiune strategică pentru civilizaţia om-maşină de la începutul secolului douăzeci şi doi.

În rândurile cosmologilor persistă o controversă pe tema sfârşitului lumii în flăcări (Big Crunch, care să corespundă începutului, Big Bang) sau în gheaţă (moartea stelelor, pe măsură ce se depărtează într-o permanentă expansiune), dar nu se ia în calcul forţa inteligenţei, ca şi cum apariţia ei n-ar fi fost decât un spectacol de divertisment colateral faţă de marea mecanică cerească, stăpâna actuală a Universului. De cât timp vom avea nevoie pentru a ne răspândi inteligenţa în forma ei nonbiologică în întregul Univers? Dacă putem depăşi viteza luminii – recunosc, un uriaş „dacă" –, de exemplu, folosind găurile de vierme din spaţiu (care sunt compatibile cu modul actual în care înţelegem noi fizica), asta s-ar putea realiza în decurs de câteva secole. Altminteri, e posibil să dureze cu mult mai mult. Indiferent de scenariu, destinul nostru este trezirea Universului şi apoi deciderea conştientă a sorţii sale insuflându-i inteligenţa noastră umană în forma ei nonbiologică.

NOTE

Introducere

1. Iată o frază din *Un veac de singurătate* de Gabriel García Márquez: „Aureliano Segundo nu băgă de seamă la această poliloghie de reproşuri până a doua zi după micul dejun când se simţi sâcâit de un bâzâit care devenise deja mai fluid şi mai zgomotos decât gălăgia ploii şi era Fernanda care umbla prin casă, lamentându-se că fusese crescută să fie regină şi ajunsese slujnică într-o casă de nebuni, cu un soţ puturos, idolatru şi libertin care aştepta să-i pice pâinea în gură din cer în timp ce ea se spetea să ţină, cu chiu, cu vai pe linia de plutire o casă în care erau atât de multe de făcut, atâtea de îndurat şi atâtea de reparat de când făcea Bunul Dumnezeu lumină pe pământ în zori şi până venea vremea de culcare, că atunci când se

întindea o usturau ochii de parcă ar fi avut sticlă pisată sub pleoape şi nimeni nu-i zicea şi ei Bună dimineaţa, *Fernandina*, *ai dormit bine*? şi nici n-o întreba careva, măcar de politeţe, de ce era atât de palidă şi de ce se scula cu cearcănele – alea violete sub ochi, deşi se aşteptase la astea, sigur că da, din partea unei familii care o considerase întotdeauna o belea, o zdreanţă, o paiaţă desenată pe perete şi care peste tot pe unde se nimerea o tot bârfea pe la spate, o făcea şoarece de biserică, farisee, vicleană, ba chiar şi Amaranta, odihnească-se în pace, zisese în gura mare că era una de-aia care-şi încurcă ţoalele cu poalele, ferească Dumnezeu, ce cuvinte! iar ea îndurase totul resemnată mulţumită Sfântului Părinte, dar nu se mai putuse abţine când diavolul ăla de José Arcadio Segundo zisese că peste familie căzuse blestemul când îşi deschisese porţile unei muntence, auzi vorbă, unei muntence încrezute, mântuieşte-ne Doamne, odraslă de muntean al dracu' din aceeaşi stirpe cu muntenii trimişi de guvern să îi omoare pe muncitori, auzi dumneata, şi nu la alta bătea, ci chiar la ea, fina Ducelui de Alba, o doamnă de asemenea descendenţă că le dădea crize de ficat nevestelor de preşedinţi, o femeie de sânge nobil ca ea, care avea dreptul să se semneze cu unsprezece patronime peninsulare şi care era singura fiinţă muritoare din orăşelul acela plin de ticăloşi care nu era pusă deloc în încurcătură când vedea şaisprezece feluri de tacâmuri în faţa ochilor, ca să aibă apoi desfrânatul de bărbat – se de ce să se prăpădească de râs şi să spună că atâtea cuţite şi furculiţe nu erau menite unei fiinţe omeneşti, ci unui miriapod şi tot ea era şi singura care putea să spună cu ochii închişi când să se servească vinul alb, pe ce parte şi în care pahar şi când vinul roşu şi din ce parte şi în ce pahar, nu ca ţăranca aia de Amaranta, odihnească-se în pace, care credea că vinul alb se serveşte ziua şi vinul roşu noaptea şi era singura de pe toată coasta care se putea mândri că se îngrijea de nevoile ei trupeşti numai în oale de noapte din aur, aşa că şi colonelul Aureliano Buendía,

odihnească-se în pace, să aibă îndrăzneala să o întrebe, cu umorul lui rău de mason, de unde îşi căpătase privilegiul şi dacă de răhăţit se răhăţea tot rahat sau nu cumva de-a dreptul busuioc, auzi vorbă, chiar în cuvintele astea, şi tocmai Renata, fiica ei, care dintr-o neatenţie o văzuse făcându-şi treaba mare în dormitor, să-i răspundă că chiar dacă oala era din aur şi cu însemne heraldice, ce era înăuntru tot rahat era, rahat de-a dreptul şi mai rău chiar decât oricare altul pentru că era rahat vârtos de munteancă, auzi vorbă de la propria ei fiică, aşa că nu-şi mai făcuse niciun fel de iluzii în privinţa restului familiei, dar, în orice caz, avea dreptul să se aştepte la ceva mai multă consideraţiune din partea soţului pentru că, de bine, de rău, era soţul ei cu cununie, consortul ei, sprijinul, defloratorul ei legal, care luase asupra lui de bunăvoie şi nesilit de nimeni grava răspundere să o ducă din casa părintească, unde nu-i lipsise şi nu se plânsese niciodată de nimic, unde împletea cununi funerare să-şi mai treacă timpul, pentru că naşul ei trimisese o scrisoare care îi purta semnătura şi pecetea de pe inelul lui în sigiliul din ceară ca să spună că mâinile finei lui nu erau făcute pentru corvezi pe lumea asta altele decât cântatul la clavir şi, cu toate astea, nebunul de soţ o luase din casa ei cu tot felul de dojeni şi atenţionări şi 6 adusese în tigaia infernală în care nu putea omu' să respire de zăduf şi până să-şi încheie ea postul de Rusalii, îşi şi luase tălpăşiţa doar în nădragi şi cu acordeonul ăla al lui de chefliu să se destrăbălleze în adulter cu o mizerabilă pe care era de-ajuns să o vezi din spate, aşa se spunea, cel puţin, era de-ajuns să-i vezi fundu' de iapă ca să-ţi dai seama că era o de-aia, o de-aia… de-a dreptul invers decât ea, care era o doamnă şi la castel şi în cocină, şi la masă şi în pat, doamnă din născare, cu frica lui Dumnezeu, care îi asculta poruncile şi se supunea dorinţelor Lui şi cu care sigur că el nu putea să-şi facă toate mendrele şi porcăriile pe care le făcea cu cealaltă care, sigur, era dispusă la orice, ca matroanele franţuzoaice şi mai rea decât ele dacă stai să

te gândeşti, pentru că alea cel puţin aveau obrazul să-şi pună felinar roşu la poartă, asemenea porcării, auzi dumneata şi numai asta îi mai lipsea unicei şi iubitei fiice a Doriei Renata Argote şi a lui Don Fernando del Carpio, mai ales a acestuia din urmă, un om cinstit, bun creştin, cavaler al Ordinului Sfântului Mormânt, unul din aceia care primiseră de-a dreptul de la Dumnezeu privilegiul să nu putrezească în groapă, cu pielea curată şi strălucitoare ca obrajii miresei şi limpede ca smaraldul".

2. Vezi graficul Creşterea volumului de date despre secvenţele ADN în Genbank de la Capitolul 10.

3. Cheng Zhang şi Jianpeng Ma, „Enhanced Sampling and Applications in Protein Folding în Explicit Solvent", *Journal of Chemical Physics* 132, nr. 24 (2010): 244101. Vezi de asemenea http://folding. stanford.edu/English/ Aboutaboutthe Foldinghomeproject, care a folosit un număr de peste cinci milioane de computere din întreaga lume pentru simularea conformării proteinelor.

4. Pentru o descriere mai amplă a acestei argumentaţii, vezi secţiunea „[The Impact...] on the Intelligent Destiny of the Cosmos: Why We Are Probably Alone in the Universe" din Capitolul 6 al cărţii *The Singularity Is Ne-ar* de Ray Kurzweil (New York: Viking, 2005).

5. James D. Watson, *Discovering the Brain* (Washington, DC: National Academies Press, 1992).

6. Sebastián Seung, *Connectome*: *How the Brain's Wiring Makes Us Who We Are* (New York: Houghton Mifflin Harcourt, 2012).

7. „Mandelbrot Zoom", http://www.youtube.com/watch?v=gew8 xpblaRA; „Fractal Zoom Mandelbrot Corner",http://www.youtube.com/ watch?v=GGBwuYuOOS.

Capitolul 1. Experimente mintale despre lume

1. Charles Darwin, *Originea speciilor* (P.F. Collier & Son, 1909), 185/95 – 96.

2. Darwin, *Originea speciilor*, 751 (206.1.1 – 6), ediţia

Peckham's Variorum, editată de Morse Peckham, *The Origin of Species by Charles Darwin*: *A Variorum Text* (Philadelphia: University of Pennsylvania Press, 1959).

3. R. Dahm, „Discovering DNA: Friedrich Miescher and the Early Years of Nucleic Acid Research", *Human Genetics* 122, nr. 6 (2008): pp. 565 – 81, doi: 10.1007/S00439 – 007 – 0433-O; PMID17901982.

4. Valéry N. Soyfer, „The Consequences of Political Dictatorship for Russian Science", *Nature Reviews Genetics* 2, nr. 9 (2001): pp. 723 – 29, doi: 10.1038/35088598; PMID11533721.

5. J.D. Watson şi F.H.C. Crick, „A Structure for Deoxyribose Nucleic Acid", *Nature* 171 (1953): 737 – 38, http://www.nature.com/nature/dna50/watsoncrick.pdf şi „Double Helix: 50 Years of. DNA", *Nature Archive*, http://www.nature.com /nature/ dna50/archive.html.

6. Franklin a murit în anul 1958, iar Premiul Nobel pentru descoperirea ADN-ului a fost conferit. În anul 1962. Există o controversă pe tema împărţirii acestui premiu, dacă Franklin ar mai fi trăit în 1962.

7. Albert Einstein, „On the Electrodynamics of Moving Bodies" (1905). Această lucrare a consacrat Teoria specială a relativităţii. Vezi Robert Bruce Lindsay şi Henry Margenau, *Foundations of Physics* (Woodbridge, CT: Ox Bow Press, 1981), pag. 330.

8. „Crookes radiometer", Wikipedia, http://en.wikipedia.org/wiki/ Crookesradiometer.

9. Observă că o parte din impulsul fotonilor este transferată moleculelor de aer din incintă (deoarece nu conţine un vid perfect) şi apoi e transmisă de către moleculele de aer încălzite palelor moriştii.

10. Albert Einstein, „Does the Inerţia of a Body Depend Upon Its Energy Content?" (1905). Acecastă lucrare a consacrat celebra formulă a lui Einstein, $E = mc^2$.

11. „Albert Einstein's Letters to President Franklin Delano Roosevelt",

http://hypertextbook.com/eworld/einstein. shtml.

Capitolul 3. Un model de neocortex:
Teoria minţii bazate pe recunoaştere

1. Unele animale nemamifere, cum ar fi ciorile, papagalii şi caracatiţele, sunt creditate cu un anumit nivel de raţiune, însă una limitată şi incapabilă să creeze unelte care să aibă propriul curs evolutiv de dezvoltare. Este posibil ca aceste animale să-şi fi adaptat alte regiuni ale creierului pentru realizarea unui număr mic de nivele de gândire ierarhică, însă, pentru gândirea relativ neîngrădită, ierarhică de care e capabil omul este necesar un neocortex.

2. V.B. Mountcastle, „An Organizing Principle for Cerebral Function: The Unit Model and the Distributed Sistem" (1978), în Gerald M. Edelman şi Vernon B; Mountcastle, *The Mindful Brain: Cortical Organization and the Group-Selective Theory of Higher Brain Function* (Cambridge, MA: MIT Press, 1982).

3. Herbert A. Simon, „The Organization of Complex Sistems", în Howard H. Pattee, ed., *Hierarchy Theory*: *The Challenge of Complex Sistems* (New York: George Braziller, Inc., 1973), http://blog. santafe.edu/wp-content/uploads/2009/03/ simon 1973.pdf.

4. Marc D. Hauser, Noam Chomsky şi W. Tecumseh Fitch, „The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?", *Science* 298 (noiembrie 2002): 1569 – 79, http://www.sciencemag.org/content/298/5598/1569. short.

5. Următorul pasaj din cartea *Transcend*: *Nine Steps to Living Well Forever*, de Ray Kurzweil şi Terry Grossman (New York: Rodale, 2009), descrie mai detaliat această tehnică de vis lucid: am pus la punct o metodă de rezolvare a problemelor în timpul somnului. Am perfecţionat-o pentru mine pe parcursul mai multor decenii şi am descoperit mijloacele subtile prin care este posibil să meargă mai bine. Încep dându-mi o problemă când mă aşez în pat. Poate să fie orice fel de problemă. Poate fi una

de matematică, un lucru care să aibă legătură cu vreuna dintre invenţiile mele, o chestiune de strategie în afaceri sau chiar o problemă interpersonal. Mă gândesc la problemă vreo câteva minute, dar nu încerc s-o rezolv. Asta nu m-ar face decât să îndepărtez soluţionarea creativă a problemei, care urmează să-mi vină. Încerc să mă gândesc la ea. Ce ştiu despre problemă? Ce formă ar putea să îmbrace soluţia? Şi pe urmă adorm. Procedând astfel, îmi pregătesc mintea inconştientă să lucreze la problemă.

Terry: Sigmund Freud a subliniat că, atunci când visăm, mulţi dintre cenzorii din creierul nostru sunt relaxaţi, aşa că putem visa lucruri care sunt tabuuri din punct de vedere social, cultural şi chiar sexual. Putem visa lucruri ciudate, la care nu ne-am permite să ne gândim în timpul zilei. Este cel puţin unul dintre motivele pentru care visele mele sunt ciudate.

Ray: Mai sunt şi jaluzele profesionale care împiedică omul să gândească creativ, multe dintre ele provenind din pregătirea noastră profesională, blocaje mentale cum ar fi „Nu poţi rezolva problema procesării semnalelor în acest fel” sau „Lingvistica nu este obligată să respecte asemenea reguli”. Aceste premise mentale se relaxează şi ele în timpul somnului, aşa că o să visez noi moduri de rezolvare a problemelor fără să mă împovărez cu aceste constrângeri din starea de veghe.

Terry: Mai există o parte a creierului care nu munceşte când dormim: facultăţile raţionale care evaluează faptul că o idee este rezonabilă sau ba. Deci acesta este un alt motiv pentru care în vise se întâmplă lucruri ciudate sau fantastice. Când trece un elefant printr-un perete, nu suntem şocaţi de posibilitatea unui asemenea lucru. Pur şi simplu ne spunem nouă înşine, celor din vis, „în regulă, a trecut un elefant printr-un perete, ce mare lucru”. Câteodată, dacă mă trezesc în puterea nopţii, constat că am visat într-un fel ciudat sau neobişnuit problema pe care mi-am dat-o spre rezolvare.

Ray: Pasul următor are loc dimineaţa, în starea

indecisă dintre somn şi veghe, adeseori numită visare lucidă. În această stare, încă mai păstrez senzaţiile şi imaginile din vis, dar sunt în posesia facultăţilor raţionale. Îmi dau seama, de exemplu, că mă aflu într-un pat. Şi pot să-mi formulez gândul că am o groază de lucruri de făcut şi că ar fi mai bine să mă scol. Însă asta ar fi o greşeală. De câte ori pot, rămân în pat şi rămân în continuare în această stare de visare lucidă, deoarece ea este cheia acestei metode creative de rezolvare a problemelor. Apropo, metoda nu funcţionează dacă sună deşteptătorul.

Cititorul: Pare să fie ce-i mai bun în cele două lumi.

Ray: Exact. Mai am încă acces la gândurile din vis legate de problema pe care mi-am dat-o cu o seară înainte, însă acum sunt suficient de conştient şi raţional pentru a evalua noile idei creative care mi-au venit în minte peste noapte. Pot să le decelez pe Cele logice. După 20 de minute, poate, petrecute în această stare, în mod invariabil am o perspectivă mult mai pătrunzătoare a problemei.

Am făcut şi invenţii în acest fel (şi mi-am petrecut restul zilei scriind cereri de patente), mi-am dat seama cum să organizez materialul pentru o carte cum este aceasta şi mi-au venit idei utile pentru un set variat de probleme. Dacă am de luat o decizie-cheie, întotdeauna aplic acest proces, după care este foarte posibil să am cu adevărat încredere în deciziile mele.

Cheia procesului este să-ţi laşi mintea liberă, fără prejudecăţi şi să nu-ţi faci griji în legătură cu felul cum funcţionează metoda. E opusul disciplinei mentale. Gândeşte-te la problemă, dar pe urmă lasă ideile să treacă peste tine când adormi. Dimineaţa, lasă-ţi din nou mintea liberă când treci 7 n revistă ideile ciudate pe care le-au generat visele. Am constatat că e o metodă nepreţuită de punere la muncă a creativităţii naturale a viselor.

Cititor: Păi, noi, ăştia, obsedaţii de muncă, acum putem să muncim şi în vis. Nu sunt sigur că soţia o să fie prea bucuroasă.

Ray: Practic, poţi să consideri că e ca şi cum ţi-ai

pune visele să muncească în locul tău.

Capitolul 4. Neocortexul biologic

1. Steven Pinker, *How the Mind Works* (New York: Norton, 1997), pp. 152 – 53.

2. D.O. Hebb, *The Organization of Behavior* (New York: John Wiley & Sons, 1949).

3. Henry Markram şi Rodrigo Perrin, „Innate Neural Assemblies for Lego Memory", *Frontiers în Neural Circuits* 5, nr. 6 (2011).

4. Comunicare e-mail de la Henry Markram, 19 februarie, 2012.

5. Van J. Wedeen şi colab., „The Geometric Structure of the Brain Fiber Pathways", *Science* 335, nr. 6076 (30 martie, 2012).

6. Tai Sing Lee, „Computations in the Early Visual Cortex", *Journal of Physiology*, Paris 97 (2003): pp. 121 – 39.

7. O listă de lucrări poate fi găsită la adresa: http://cbel. mit.edu/people/pogglo/ tpevshortpubs.pdf.

8. Daniel J. Felleman şi David C. Van Essen, „Distributed Hierarchical Processing in the Primate Cerebral Cortex", *Cerebral Cortex* 1, nr. 1 (ianuarie/februarie 1991): 1 – 47. O analiză convingătoare a matematicii bayesiene a comunicaţiilor de sus în jos şi de jos în sus la nivelul neocortexului e oferită de Tai Sing Lee în „Hierarchical Bayesian Inference in the Visual Cortex", *Journal of the Optical Society of America* 20, nr. 7 (iulie 2003): pp. 1434 – 48.

9. Un Hasson şi colab., „A Hierarchy of Temporal Receptive Windows în Human Cortex", *Journal of Neuroscience* 28, nr. 10 (5 martie, 2008): pp. 2539 – 50.

10. Marina Bedny şi colab., „Language Processing in the Occipital Cortex of Congenitally Blând Adults", *Proceedings of the National Academy of Sciences* 108, nr. 11. (15 martie, 2011): pp. 4429 – 34.

11. Daniel E. Feldman, „Synaptic Mechanisms for Plasticity în Neocortex", *Annual Review of Neuroscience*

32 (2009): pp. 33 – 55.

12. Aaron C. Koralek şi colab., „Corticostriatal Plasticity Is Necessary for Learning Intenţional Neuroprosthetic Skills”, *Nature* 483 (15 martie, 2012): pp. 331 – 35.

13. Comunicare prin e-mail de la Randal Koene, ianuarie 2012.

14. Min Fu, Xinzhu Yu, Ju Lu şi Yi Zuo, „Repetitive Motor Learning Induces Coordinated Formation of Clustered Dendritic Spines in Vivo”, *Nature* 483 (1 martie, 2012): pp. 92 – 95.

15. Dario Bonanomi şi colab., „Ret Is a Multifuncţional Coreceptor That Integrates Diffusible – and Contact – Axon Guidance Signals”, *Cell* 148, nr. 3 (februarie 2012): pp. 568 – 82.

16. Vezi nota de final 7 de la Capitolul 11.

Capitolul 5. Creierul vechi

1. Vernon B. Mountcastle, „The View from Within: Pathways to the Study of Perception”, *John's Hopkins Medical Journal* 136 (1975): pp. 109 – 31.

2. B. Roska şi F. Werblin, „Vertical Interactions Across Ten Paralel, Stacked Representations in the Mammalian Retina”, *Nature* 410, nr. 6828 (29 martie, 2001): pp. 583 – 87; „Eye Strips Images of All but Bare Essentials Before Sending Visual Information to Brain, UC Berkeley Research Shows”, comunicat de presă al Universităţii California, Berkeley, 28 martie 2001, www.berkeley.edu/news/media/ releases/2001/03/28 wers1.html.

3. Lloyd Watts, „Reverse-Engineering the Human Auditory Pathway”, în J. Liu şi colab., eds., WCC12012 (Berlin: Springer-Verlag, 2012), pp. 47 – 59. Lloyd Watts, „Real-Time, High-Resolution Simulation of the Auditory Pathway, with Application to Cell-Phone Noise Reduction”, *ISCAS* (2 iunie, 2010): pp. 3821 – 24. Pentru alte lucrări, vezi http://www.lloydwatts.com/publications.html.

4. Vezi Sandra Blakeslee, „Humanity? Maybe It's All

in the Wiring", *New York Times*, 11 decembrie 2003,http://www.nytimes.com/2003/12/09/science/09BRAI.html.

5. T.E.J. Behrens şi colab., „Non-invasive Mapping of Connections between Human Thalamus and Cortex Using Diffusion Imaging", *Nature Neuroscience* 6, nr. 7 (iulie 2003): pp. 750 – 57.

6. Timothy J. Buschman şi colab., „Neural Substrates of Cognitive Capacity Limitations", *Proceedings of the National Academy of Sciences* 108, nr. 27 (5 iulie 2011): pp. 11252 – 55, http://www.pnas.Org/content/108/27/11252. long.

7. Theodore W. Berger şi colab., „A Cortical Neural Prosthesis for Restoring and Enhancing Memory", *Journal of Neural Engineering* 8, nr. 4 (august 2011).

8. Funcţiile de bază sunt funcţii neliniare care pot fi combinate liniar (adunându-se funcţii de bază cu greutăţi diferite) pentru aproximarea oricărei funcţii neliniare. A. Pouget şi L.H. Snyder, „Computational Approaches to Senzorimotor Transformations", *Nature Neuroscience* 3, nr. 11 *supliment* (noiembrie 2000): pp. 1192 – 98.

9. J.R. Bloedel, „Funcţional Heterogeneity with Structural Homogeneity: How Does the Cerebellum Operate?", *Behavioral and Brain Sciences* 15, nr. 4 (1992): pp. 666 – 78.

10. S. Grossberg şi R.W. Paine, „A Neural Model of Cortico-Cerebeliar Interactions during Attentive Imitation and Predictive Learning of Sequential Handwriting Movements", *Neural Networks* 13, nr. 8 – 9 (octombrie-noiembrie 2000): pp. 999 – 1046.

11. Javier F. Medina şi Michael D. Mauk, „Computer Simulation of Cerebellar Information Processing", *Nature Neuroscience* 3 (noiembrie 2000): pp. 1205 – 11.

12. James Olds, „Pleasure Centers in the Brain", *Scientific American* (octombrie 1956): pp. 105 – 16. Aryeh Routtenberg, „The Reward Sistem of the Brain", *Scientific American* 239 (noiembrie 1978): pp. 154 – 64. K.C.

Berridge şi M.L. Kringelbach, „Affective Neuroscience of Pleasure: Reward în Humans and Other Animals”, *Psychopharmacology* 199 (2008): pp. 457 – 80. Morten L. Kringelbach, *The Pleasure Center*: *Trust Your Animal Instincts* (New York: Oxford University Press, 2009). Michael R. Liebowitz, *The Chemistry of Love* (Boston: Little, Brown, 1983). W.L. Witters şi P. Jones-Witters, *Human Sexuality*: *A Biological Perspective* (New York: Van Nostrand, 1980).

Capitolul 6. Abilităţile transcendentale

1. Michael Nielsen, *Reinventing Discovery*: *The New Era of Networked Science* (Princeton, NJ: Princeton University Press, 2012), 1 – 3. T. Gowers şi M. Nielsen, „Massively Collaborative Mathematics”, *Nature* 461, nr. 7266 (2009): pp. 879 – 81. „A Combinatorial Approach to Density Hales-Jewett”, *Gowers's Weblog*, http://gowers.wordpress.com/2009/02/01/a-combinatorial-approach-todensity-halesjewett/. Michael Nielsen, „The Polymath Project: Scope of Participation”, 20 martie 2009, http://michaelnielsen.org/blog/? p = 584. Julie Rehmeyer, „SIAM: Massively Collaborative Mathematics”, *Society for Industrial and Applied Mathematics*, 1 aprilie 2010, http://www.siam.org/news/news.php? id = 1731.

2. P. Dayan şi Q.J.M. Huys, „Serotonin, Inhibition, and Negative Mood”, *PLOS Computational Biology* 4, nr. 1 (2008), http://compbiol. plosjournals.org/ periserv/? request = get-document & doi = 10.1371/joumal. pebi.0040004.

Capitolul 7. Neocortexul digital inspirat de biologie

1. Gary Cziko, *Without Miracles*: *Universal Selection Theory and the Second Darwinian Revolution* (Cambridge, MA: MIT Press, 1955).

2. David Dalrymple; am fost mentorul lui de când avea opt ani, în anul 1999. Poţi să-i citeşti biografia aici: http://esp. mit.edu/learn/teachers/davidad/bio.html şi http://www.brainsciences.org/Research-Team/mi-david-

dalrymple.html.

3. Jonathan Fildeş, „Artificial Brain «10 Years Away»”, BBC News, 22 iulie 2009, http://news. bbe.co. uW2/hi/8164060. stm. Vezi de asemenea materialul video „Henry Markram on Simulating the Brain: The Next Decisive Years”, http://www.kurzweilai.net/henry-markram-simulating-the-brain-next-decisive-years.

4. M. Mitchell Waldrop, „Computer Modelling: Brain in a Box”, *Nature News*, 22 februarie 2012, http://www.nature.com/news computer-modelling-brainin-a-box-l.10066.

5. Jonah Lehrer, „Can a Thinking, Remembering, Decision-Making Biologically Accurate Brain Be Built from a Supercomputer?”, *Seed*, http://seedmagazine.com/ content/article/out of the blue/.

6. Fildeş, Artificial Brain «10 Years Away»”.

7. Vezi http://www.humanconnectoriieproject.org/.

8. Anders Sandberg şi Nick Bostrom, *Whole Brain Emulation*: *A Roadmap*, *Tehnical Reporte 2008 – 3 (2008)*, Future of Humanity Institute, Oxford University, www.fhi. ox.ac.uk/reports/2008 – 3.pdf.

9. Iată schema de bază pentru un algoritm de reţea neuronală. Sunt posibile numeroase variaţiuni, iar creatorul sistemului trebuie să includă anumiţi parametri şi metode critice detaliate în paginile care urmează.

Crearea unei soluţii de reţea neuronală pentru o problemă presupune următorii paşi:

Definirea inputului.

Definirea topologiei reţelei neuronale (respectiv, straturile de neuroni şi conexiunile dintre neuroni).

Pregătirea reţelei neuronale cu exemple ale problemei.

Rularea reţelei neuronale pregătite pentru rezolvarea unor noi exemple ale problemei.

Lansarea pentru public a reţelei neuronale.

Aceşti paşi (cu excepţia ultimului) sunt detaliaţi mai jos:

Inputul problemei inputul problemei în reţeaua neuronală rezidă într-o serie de numere. Acest imput poate fi:

Într-un sistem de recunoaştere vizuală a modelelor, un şir bidimensional de numere reprezentând pixelii dintr-o imagine; sau într-un sistem de recunoaştere auditiv (de exemplu, a vorbirii), un şir bidimensional de numere reprezentând un sunet, prima dimensiune reprezentând parametrii sunetului (de exemplu componentele frecvenţei), iar a doua dimensiune – diverse momente în timp; sau Un sistem arbitrar de recunoaştere a modelelor, un şir de numere reprezentând modelul de imput.

Definirea topologiei

Pentru crearea unei reţele neuronale, arhitectura fiecărui neuron constă în:

Imput-uri multiple, cu fiecare Imput „conectat" fie la outputul altui neuron, fie la unul dintre numerele din Imput.

În general, un singur output, conectat fie la Imputul altui neuron (care se află de regulă pe un strat superior), fie la outputul final.

Setarea primului strat de neuroni

Se creează N0 neuroni pe primul strat. Toţi aceşti neuroni sunt „conectaţi" cu toate căile lor de Imput la „puncte" (respectiv, numere) din inputul problemei. Aceste conexiuni pot fi determinate aleatoriu sau folosindu-se un algoritm evolutiv (vezi în continuare).

Se atribuie o „tărie sinaptică" fiecărei conexiuni create. Aceste tării pot să pornească de la acelaşi nivel, pot fi atribuite aleatoriu sau pot fi determinate într-un alt mod (vezi în continuare).

Setarea straturilor suplimentare de neuroni

Se setează un total de M straturi de neuroni. Pentru fiecare strat, se setează neuronii de pe strat.

Pentru stratul /:

Se creează N (neuroni pe stratul /. Se „conectează" fiecare dintre inputurile multiple ale fiecărui neuron la

output-urile neuronilor de pe stratul 1 (vezi variaţiunile care urmează).

Se atribuie o „tărie sinaptică" fiecărei conexiuni create. Aceste tării pot să pornească de la acelaşi nivel, pot fi atribuite aleatoriu sau pot fi determinate într-un alt mod (vezi în continuare).

Output-urile neuronilor de pe stratul M sunt output-urile reţelei neuronale (vezi variaţiunile în continuare).

Încercările de recunoaştere

Cum funcţionează fiecare neuron

Odată reţeaua setată, ea face următoarele pentru fiecare încercare de recunoaştere:

Fiecare imput al neuronului care corespunde tăriei este computat prin multiplicarea output-ului altui neuron (sau a inputului iniţial) la care este conectat inputul acestui neuron, după tăria sinaptică a acelei conexiuni. Toate inputurile neuronului care corespund ca tărie sunt însumate.

Dacă această sumă depăşeşte pragul de excitare al acestui neuron, se consideră că neuronul emite, iar outputul lui este 1. Altminteri, outputul lui este O (vezi variaţiunile în continuare).

Se fac următoarele pentru fiecare încercare de recunoaştere:

Pentru fiecare strat, de la stratul o la stratul *M*:

Pentru fiecare neuron de pe strat:

Se însumează inputurile care corespund ca tărie (fiecare imput ce corespunde ca tărie = outputul altui neuron [sau inputul original] la care este conectat inputul acestui neuron, înmulţit cu tăria sinaptică a conexiunii respective).

Dacă suma de inputuri corespunzătoare tăriei este mai mare decât pragul de excitaţie al acestui neuron, se setează outputul neuronului la 1; altminteri, se setează la O.

Pentru pregătirea reţelei neuronale:

Se rulează încercări repetate de recunoaştere pe

probleme-mostră. După fiecare încercare, se ajustează tăria sinaptică a tuturor conexiunilor interneuronale pentru îmbunătăţirea performanţelor reţelei neuronale la această încercare (vezi discuţia despre cum se procedează în continuare).

Se continuă această pregătire până când acurateţea reţelei neuronale nu se mai îmbunătăţeşte (respectiv, devine asimptotică).

Decizii-cheie de concepţie:

În schema simplă de mai înainte, creatorul algoritmului reţelei neuronale trebuie să determine de la bun început:

Ce reprezintă numerele de imput.

Numărul de straturi de neuroni.

Numărul de neuroni de pe fiecare strat. (Nu este neapărat necesar ca toate straturile să aibă acelaşi număr de neuroni.)

Numărul de inputuri pentru fiecare neuron de pe fiecare strat. Numărul de inputuri (respectiv, de conexiuni interneuronale) poate şi el să varieze de la un neuron la altul şi de la un strat la altul.

„Cablarea" propriu-zisă (respectiv, realizarea conexiunilor). Pentru fiecare neuron de pe fiecare strat, aceasta constă într-o listă de alţi neuroni, ale căror output-uri constituie inputuri pentru neuronul respectiv. Este vorba despre o parte-cheie a concepţiei. Există mai multe căi posibile prin care se poate face acest lucru:

(1) Reţeaua neuronală este cablată aleatoriu; ori

(2) Se foloseşte un algoritm evolutiv (vezi în continuare) pentru a se determina cablarea optimă; ori

(3) Se foloseşte judecata creatorului pentru determinarea cablării.

Tăriile sinaptice iniţiale (respectiv, greutatea) fiecărei conexiuni. Există mai multe moduri posibile de realizare:

(1) Se setează tăriile sinaptice la aceeaşi valoare; ori

(2) Se setează tăriile sinaptice la valori aleatorii diferite; ori

(3) Se foloseşte un algoritm evolutiv pentru determinarea unui set optim de valori iniţiale; orí

(4) Se foloseşte judecata creatorului pentru determinarea valorilor iniţiale. Pragul de excitaţie al fiecărui neuron.

Se determină outputul. Outputul poate să fie:

(1) Output-urile stratului *M* de neuroni; ori

(2) Outputul unui singur neuron, ale cărui imput-urí sunt output-urile neuronilor de pe stratul M; ori

(3) o funcţie (de exemplu, suma) output-urilor neuronilor de pe stratul M; ori

(4) o altă funcţie a output-urilor neuronilor de pe mai multe straturi.

Se determină felul cum sunt ajustate tăriile sinaptice ale tuturor conexiunilor în cursul pregătirii acestei reţele neuronale. Este o decizie-cheie de concepţie şi e subiectul unui volum impresionant de cercetări şi discuţii. Există mai multe căi posibile prin care se poate face acest lucru:

(1) La fiecare încercare de recunoaştere, se măreşte sau se reduce fiecare tărie sinaptică cu o valoare fixă (în general mică), astfel încât outputul reţelei neuronale să corespundă mai bine răspunsului corect. Un mijloc este să se încerce şi creşterea, şi descreşterea şi să se vadă care are efectul cel mai dezirabil. Este o metodă cronofagă, astfel încât există şi alte metode pentru luarea deciziilor în sensul creşterii sau descreşterii fiecărei tării sinaptice.

(2) Există şi alte metode statistice pentru modificarea tăriilor sinaptice după fiecare încercare, astfel încât performanţa reţelei neuronale la încercarea respectivă să se apropie mai mult de răspunsul corect.

Ase reţine că pregătirea reţelei neuronale dă rezultate chiar şi dacă încercările de pregătire nu sunt toate corecte. Acest fapt permite folosirea de date de pregătire din lumea reală, care pot avea o rată de erorí inerentă. Una dintre cheile succesului unui sistem de recunoaştere bazat pe reţele neuronale este volumul de date folosit la pregătire. De obicei, e necesar un volum

foarte substanţial pentru obţinerea unor rezultate satisfăcătoare. La fel ca în cazul studenţilor umani, volumul de timp pe care îl petrece o reţea neuronală în învăţarea lecţiilor este un factor-cheie pentru performanţa lui.

Variapuni

Sunt fezabile mai multe variaţiuni ale celor de mai înainte. De exemplu:

Există mai multe moduri de determinare a topologiei. În particular, cablarea interneuronală poate fi setată fie aleatoriu, fie folosindu-se un algoritm evolutiv. Există mai multe mijloace de setare a tăriilor sinaptice iniţiale.

Imput-urile pentru neuronii de pe stratul / nu trebuie să vină neapărat din output-urile neuronilor de pe stratul /*1*. Invers, inputurile pentru neuronii de pe fiecare strat pot să vină de la orice strat inferior sau de la orice strat în mod general.

Există mai multe moduri de determinare a output-ului final.

Metoda descrisă mai înainte se concretizează prin excitaţii „Totul sau nimic" (1 ori O), situaţie denumită nonliniaritate. Mai există şi alte funcţii nonliniare care pot fi folosite. În mod curent, se foloseşte o funcţie care ajunge de la 0 la 1 de o manieră mai rapidă dar mai treptată. De asemenea, output-urile pot fi şi alte numere, nu doar o şi 1.

Diversele metode de ajustare a tăriilor sinaptice în cursul pregătirii constituie decizii-cheie de concepţie.

Schema de mai înainte reprezintă o reţea neuronală „sincronă" în care fiece recunoaştere de model se realizează prin computarea outputurilor de la fiecare nivel, începând cu stratul o şi mergând până la stratul *M*. în cazul unui sistem paralel real, în care fiece neuron operează independent de ceilalţi, neuronii pot opera „asincron" (respectiv, independent). În cazul unei abordări asincrone, fiecare neuron îşi scanează permanent inputurile şi emite de câte ori suma inputurilor sale ce corespund tăriei îi depăşeşte pragul (sau orice altceva ar

specifica funcţia sa de output).

10. Robert Mannell, „Acoustic Representations of Speech”, 2008, http://clas. mq.edu.auzacoustics/frequency/acousticspeech; html.

11. Iată schema de bază pentru un algoritm genetic (evolutiv) Sunt posibile multe variaţiuni, iar creatorul sistemului trebuie să asigure anumiţi parametri şi metode critice, detaliate în continuare.

Algoritmul evolutiv

Se creează A/” fiinţe”-soluţii. Flecare are:

Un cod genetic: o succesiune de numere ce caracterizează o anumită soluţie a problemei. Numerele pot constitui parametri critici, paşi către o soluţie, reguli etc.

Pentru fiecare generaţie de evoluţie se procedează după cum urmează:

Se procedează după cum urmează pentru fiecare dintre cele N fiinţe soluţii:

Se aplică soluţia fiinţei-soluţie (aşa cum este reprezentată de codul său genetic) la problemă ori la mediul simulat. Se evaluează soluţia.

Se aleg *L* fiinţe-soluţie cu evaluările cele mai bune care supravieţuiesc în generaţia următoare.

Se elimină (*N*-*L*) fiinţe-soluţie care nu supravieţuiesc.

Se creează (*N*-*L*) noi fiinţe-soluţie din cele L fiinţe-soluţie care au supravieţuit, prin:

(1) Realizarea de copii ale celor *L* fiinţe supravieţuitoare. Se introduc mici variaţii aleatorii pentru fiecare copie; ori

(2) Se creează alte fiinţe-soluţie suplimentare, prin combinarea unor părţi din codul genetic (folosindu-se reproducerea „sexuală” ori combinarea unor porţiuni de cromozomi) din fiinţele supravieţuitoare L; ori

(3) Se face o combinaţie din (1) şi (2).

Se decide dacă se continuă sau nu evoluţia:

Îmbunătăţire – (evaluările cele mai înalte ale generaţiei) – (evaluările cele mai înalte ale generaţiei

anterioare).

Dacă îmbunătăţirea < Pragul de îmbunătăţire, atunci am terminat. Fiinţa-soluţie cu evaluarea cea mai înaltă din ultima generaţie de evoluţie are soluţia cea mai bună. Se aplică soluţia definită de codul ei genetic la problemă.

Decizii-cheie de concepţie în schema simplă de mai înainte, creatorul trebuie să determine de la bun început:

Parametrii-cheie:

N

L

Pragul de îmbunătăţire.

Ce reprezintă numerele din codul genetic şi cum este computată soluţia din acest cod.

O metodă de determinare a celor *N* fiinţe-soluţie din prima generaţie. În general, acestea trebuie să fie doar tentative „rezonabile” de soluţie. Dacă soluţiile din prima generaţie sunt prea îndepărtate, este posibil ca algoritmul evolutiv să aibă dificultăţi în privinţa ajungerii la o soluţie bună. Adeseori, merită să se creeze de la bun început fiinţe-soluţie iniţiale astfel încât să fie rezonabil de diferite. Asta ajută procesul evolutiv să nu se blocheze la o soluţie optimă „locală”.

Cum se evaluează soluţiile.

Cum se reproduc fiinţele-soluţie supravieţuitoare.

Variaţiuni

Sunt fezabile mai multe dintre variaţiunile de mai înainte. De exemplu:

Nu trebuie să existe un număr fix de fiinţe-soluţie supravieţuitoare (L) pentru fiecare generaţie. Regula (regulile) supravieţuirii poate (pot) să permită un număr variabil de supravieţuitori.

Nu trebuie să existe un număr fix de noi fiinţe-soluţie în fiecare generaţie (*N* – *L*). Regulile de procreare pot fi independente de mărimea populaţiei.

Procrearea poate fi asociată cu supravieţuirea, prin aceasta permiţându-li-se celor mai adecvate fiinţe-soluţie să se reproducă cel mai des.

Decizia continuării sau stopării evoluţiei poate fi diversificată. Ea poate avea în vedere nu numai fiinţele-soluţie cu cele mai bune evaluări din cea (cele) mai recentă (recente) generaţie (generaţii). Poate să ia în considerare şi o tendinţă care să meargă mai în urmă de ultimele două generaţii.

12. Dileep George, „How the Brain Might Work: A Hierarchical and Temporal Model for Learning and Recognition" (dizertaţie de doctorat, Universitatea Stanford, iunie 2008).

13. A.M. Turing, „Computing Machinery and Intelligence", *Mind*, octombrie 1950.

14. Hugh Loebner are o competiţie dotată cu „Premiul Loebner" care se desfăşoară în fiecare an. Medalia de argint Loebner este acordată unui computer care trece Testul Turing original exclusiv text. Medalia de aur va fi acordată unui computer care poate să treacă o versiune a testului ce include Imput şi output audio şi video. Din punctul meu de vedere, de fapt, includerea testului audio şi video nu creşte dificultatea probei.

15. „Cognitive Assistant That Leams and Organizes", *Artificial Intelligence Center*, *SRI International*, http://www.ai. sri.com/project/CALO.

16. Dragon Go! *Nuance Communications*, *Inc.*, http://www.nuance.com/products /dragon-go-in-action/index.htm.

17. „Overcoming Artificial Stupidity", *Wolfram Alpha Blog*, 17 aprilie 2012, http://blog. wolframalpha.com/author/Stephenwolfram/.

Capitolul 8. Mintea ca un computer

1. Salomon Bochner, *A Biographical Memoir of John von Neumann* (Washington, DC: National Academy of Sciences, 1958).

2. A.M. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem", *Proceedings of the London Mathematical Society* Series 2, vol. 42 (1936 – 37): pp. 230 – 65, http://www.comlab.

ox.ac.uk/activities/ieg/e-library/ sources/tp2-le.pdf. A.M. Turing, „On Computable Numbers, with an Application to the Entscheidungsproblem: A Corection", *Proceedings of the London Mathematical Society* 43 (1938): pp. 544 – 46.

3. John von Neumann, „First Draft of a Report on the EDVAC", Moore School of Electrical Engineering, University of Pennsylvania, 30 iunie 1945. John von Neumann, „A Mathematical Theory of Communication", *Bell Sistem Tehnical Journal*, iulie şi octombrie 1948.

4. Jeremy Bernstein, *The Analytical Engine*: *Computers-Past, Present, and Future*, ediţie revăzută (New York: William Morrow & Co., 1981).

5. „Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List", *Top 50*0, 11 noiembrie 2011, http://top500.org/lists/2011/11/press-release.

6. Carver Mead, *Analog VLSI and Neural Sistems* (Reading, MA: Addison-Wesley, 1986).

7. „IBM Unveils Cognitive Computing Chips", comunicat de presă IBM, 18 august 2011, http://www-03. ibm.com/press/us/en/pressrelease/35251. wss.

8. „Japan's K Computer Tops 10 Petaflop/s to Stay Atop TOP500 List".

Capitolul 9. Experimente mintale despre minte

1. John R. Searle, „I Married a Computer", în Jay W. Richards, ed., *Are VJE Spiritual Machines*? *Ray Kurzweil* vs. *the Critics of Strong Al* (Seattle: Discovery Institute, 2002).

2. Stuart Hameroff, *Ultimate Computing*: *Biomolecular Consciousness and Nanotechnology* (Amsterdam: Elsevier Science, 1987).

3. P.S. Sebei şi colab., „The Incidence of Awareness during Anesthesia: A Multicenter United States Study", *Anesthesia and Analgesia* 99 (2004): pp. 833 – 39.

4. Stuart Sutherland, *The International Dictionary of Psychology* (New York: Macmillan, 1990).

5. David Cockbum, „Human Beings and Giant Squids", *Philosophy* 69, nr. 268 (aprilie 1994): pp. 135 –

50.

6. Ivan Petrovici Pavlov, dintr-o prelegere ţinută în anul 1913, publicată în *Lectures on Conditioned Reflexes*: *Twenty-Five Years of Obiective Study of the Higher Nervous Activity* [*Behavior*] *of Animals* (London: Martin Lawrence, 1928), pag. 222.

7. Roger W. Sperry, din James Arthur, *Lecture on the Evol*ution *of the Human Brain*, 1964, pag. 2.

8. Henry Maudsiey, „The Double Brain", *Mind* 14, nr. 54 (1889): pp. 161 – 87.

9. Susan Curtiss şi Stella de Bode, „Language after Hemispherectomy", *Brain and Cognition* 43, nr. 1 – 3 (iunie-august 2000): pp... 135 – 38.

10. E.P. Vining şi colab., „Why Would You Remove Haifa Brain? The Outcome of 58 Children after Hemispherectomy – the John's Hopkins Experience: 1968 to 1996", *Pediatrics* 100 (august 1997): pp. 163 – 71. M.B. Pulsifer şi colab., „The Cognitive Outcome of Hemispherectomy în 71 Children", *Epilepsia* 45, nr. 3 (martie 2004): pp. 243 – 54".

11. S. McClelland III şi R.E. Maxwell, „Hemispherectomy for Intractable Epilepsy în Adults: The First Reported Series", *Annals of Neurology* 61, nr. 4 (aprilie 2007): pp. 372 – 76.

12. Lars Muckli, Marcus J. Naumerd şi Wolf Singer, „Bilateral Visual Field Maps in a Patient with Only One Hemisphere", *Proceedings of the National Academy of Sciences* 106, nr. 31 (4 august 2009), http://dx.doi.org/10.1073/pnas.0809688106.

13. Marvin Minsky, *The Society of Mind* (New York: Simon and Schuster, 1988).

14. F. Fay Evans-Martin, *The Nen/ous Sistem* (New York: Chelsea House, 2005), http://www.scribd.com/doc/5012597/The-Nervous-System.

15. Benjamin Libet, *Mind Time*: *The Temporal Factor* în *Consciousness* (Cambridge, MA: Harvard University Press, 2005).

16. Daniel C. Dennett, *Freedom Evolves* (New York: Viking, 2003).

17. Michael S. Gazzaniga, *Who'se în Charge*? *Free Will and the Science of the Brain* (New York: Ecco/HarperCollins, 2011).

18. David Hume, *An Enquiry Concerning Human Understanding* (1765), ediţia a doua, sub redacţia Eric Steinberg (Indianapolis: Hackett, 1993).

19. Arthur Schopenhauer, *The Wisdom of Life*.

20. Arthur Schopenhauer, *On the Freedom of the Will* (1839).

21. Din Raymond Smullyan, *5.000 B.C. and Other Philosophical Fantasies* (New York: St. Martin's Press, 1983).

22. Pentru o examinare lămuritoare şi incitantă pe teme similare legate de identitate şi conştiinţă, vezi Martine Rothblatt, „The Terasem Mind Uploading Experiment", *International Journal of Machine Consciousness* 4, nr. 1 (2012): pp. 141 – 58. În această lucrare, Rothblatt examinează tema identităţii sub aspectul unui software care imită o persoană în temeiul unei „baze de date de interviuri video şi informaţii asociate referitoare la persoana predecesoare". În cadrul acestui viitor experiment propus, software-ul imită cu succes persoana pe care o copiază.

23. „How Do You Persist When Your Molecules Don't?", *Science and Consciousness Review* 1, nr. 1 (iunie 2004), http://www.sci-con.org/articles/20040601.html.

Capitolul 10. Legea accelerării veniturilor aplicată creierului

1. „DNA Sequencing Costs", *National Human Genome Research Institute*, NIH, http://www.genome.gov/sequencingcosts/.

2. „Genetic Sequence Data Bank, Distribution Release Notes", 15 decembrie 2009, *National Center for Biotechnology Information, National Library of Medicine*, ftp: //ftp. nebi. nih.gov/genbank/gbrel.txt.

3. „DNA Sequencing – The History of DNA Sequencing”, 2 ianuarie 2012, http://www.dnasequencing.org/history-of-dna.

4. „Cooper’s Law”, *Arraylomm*, http://www.arraycomm.com/technology/ coopers-law.

5. „The Zettabyte Era”, *Cisco*, http://www.cisco.com/en/US/solutions/collateral/ ns341/ns525/ns537/ns705/ns827/UNIH – yperconnectivityWP.html şi „Number of Internet Hosts”, *Internet Sistems Consortium*, http://www.isc.org/solutions /survey/ history.

6. Tele Geography e PriMetrica, Inc., 2012.

7. Dave Kristula, „The History of the Internet” (martie 1997, actualizat august 2001), „Internet Timeline v8. O” http://www.zakon.org/robert/internet/timeline; Quest Communications, 8-K for 9/13/1998 EX-99.1; Converge! *Network Digest*, 5 decembrie 2002, http://www.convergedigest.com/Daily/daily.asp? va = v9 n229 & fecha = December%2005, %202002; Jim Duffy, „AT & T Plâns Backbone Upgrade to 40G”, *Computerworld*, 7 iunie 2006,http://www.computerworld.com/action/article history. shtm!; Robert Zakon, „Hobbes net www.davesite.com/webstation/http://.de? command = view Articlebasic & articleld = 9001032; „40G: The Fastest Connection You Can Get?” *Internet. News. cow*, 2 noiembrie 2007, http://www.internetnews.com/infra/artide.php/3708936; „Verizon First Global Service Provider to Deploy 100G on U.S. Long-Haul Network”, comunicat de presă, *Verizon*, http://newscenter. verizon.com/press-releases/verizon/2011/verizon-first-global-service.html.

8. Facebook, „Key Facts”, http://newsroom.fb.com/content/default. aspx? Newsăreald = 22.

9. http://www.kurzweilai.net/how-my-predicţions-are-faring.

10. Calcule pe secundă ia 1.000 de dolari.

Anul Calcule pe sec. Maşină Logaritm la 1.000 de dolari natural

1900	1 -06	5.82E	Analytical Engine	— 12.05404
1908	1 -04	1.30E	Hollerith Tabulator	— 8.948746
1911	1 -05	5.79E	Monroe Calculator	— 9.757311
1919	1 -03	1.06E	IBM Tabulator	— 6.845721
1928	1 -04	6.99E	National Ellis 3.000	— 7.265431
1939	1 -03	8.55E	Zuse2	4.762175
1940	1 -02	1.43E	Bell Calculator Model	14 246797
1941	1 -02	4.63E	Zuse3	— 3.072613
1943	1 E + 00	5.31	Colossus	1.6 692151
1946	1 -01	7.98E	ENIAC	— O.225521
1948	1 JOE-01	3	IBM SSEC	— O.994793
1949	1 + 00	1.84E	BINAC	O. 6081338
1949	1 + 00	1.04E	EDSAC	0.0 430595
1951	1 + 00	1.43E	Univac 1	O. 3576744

1953	1 + 00	6.10E 1103	Univac	1.8 089443
1953	1 + 01	1.19E	IBM 701	2.4 748563
1954	1 -01	3.67E	EDVAC	— 1.00266 6
1955	1 + 01	1.65E	Whirlwind	2.8 003255
1955	1 + 00	3.44E	IBM 704	1.2 348899
1958	1 01	3.26E- 1.000	Datamatic	— 1.12177 9
1958	1 01	9.14E-	Univac II	0. 089487
1960	1 00	1.51E +	IBM 1620	O. 414755 2
1960	1 02	1.52E +	DEC PDP	— 15.0205 856
1961	1 02	2.83E +	DEC PDP	— 45.6436 786
1962	1 01	2.94E +	Univac III	3. 382014 6
1964	1 02	1.59E +	CDC 6600	5. 066385 3
1965	1 02	4.83E +	IBM 1130	6. 179188 2
1965	1 03	1.79E +	DEC PDP	— 87.4910

					876
966	1	01	4.97E+	IBM 360 Model 75	3.906407
968	1	02	2.14E+	DEC PDP-10	5.364105
973	1	02	7.29E+	Intellec8	6.591124
973	1	03	3.40E+	Data General Nova	8.131824
975	1	04	1.06E+	Altair 8800	9.266720
976	1	02	7.77E+	DEC PDP-11 Model 70	6.655440
977	1	03	3.72E+	Crayl	8.221478
977	1	04	2.69E+	Apple II	10.198766
979	1	03	1.11E+	DEC VAX 11 Model 780	7.015712
980	1	03	5.62E+	Sun-l	8.634264
982	1	05	1.27E+	IBM PC	11.748788
982	1	05	1.27E+	Compaq Portable	11.748788
983	1	04	8.63E+	IBM AT-, 80286	11.365353

1984	1 04	8.50E + 11	Apple MacIntosh	.350759		
1986	1 05	5.38E + 13	Compaq Deskpro 386	.195986		
1987	1 05	2.33E + 12	Apple Mac II	.357076		
1993	1 06	3.55E + 15	Pentium PC	.082176		
1996	1 07	4.81E + 17	Pentium PC	.688377		
1998	1 08	1.33E + 18	Pentium II PC	.708113		
1999	1 08	7.03E + 20	Pentium III PC	.370867		
2000	2 08	1.09E + 18	IBM ASCI White	.506858		
2000	2 08	3.40E + 19	MacIntosh Power G4/500	.644456		
2003	2 09	2.07E + 21	MacIntosh Power G5 2.O	.450814		
2004	2 09	3.49E + 21	Dell Dimension 8400	.973168		
2005	2 09	6.36E + 22	Power Mac G5 Quad	.573294		
2008	2 10	3.50E + 24	Dell XPS 630	.278614		
2008	2 10	2.07E + 23	MacPro	.7534		
2009	2 10	1.63E + 23	Intel Core ¡7 Desktop	.514431		
2010	2 10	5.32E + 24	Intel Core ¡7 Desktop	.697324		

11. Top 500 Supercomputer Sites, http://top500.org/.

12. „Microprocessor Quick Reference Guide”, Intel Research, http://www.intelcom/pressroom/kits/quickreffam.htm.

13.1971 – 2000: VLSI Research Inc.

2001 – 2006: *The International Technology Roadmap for Semiconductors*, 2002 Update and 2004 Update, Table 7 a, „Cost-Near-term Years”, „DRAM cost/bit at (packaged microcents) at production”.

2007 – 2008: *The International Technology Roadmap for Semiconductors*, 2007, Tables 7 a and 7 b, „Cost-Near-term Years”, „Cost-Long-term Years”, http://www.itrs.net/Links/2007ITRS/Execăum2007.pdf.

2009 – 2022: *The International Technology Roadmap for Semiconductors*, 2009, Tables 7 a and 7 b, „Cost-Near-term Years”, „Cost-Long-term Years”, http://www.itrs.net/Links/2009ITRS/Home2009.htm.

14. Pentru a face comparabilă valoarea dată? n dolari, preţurile computerelor pentru toţi anii au fost convertite în echivalentul lor din anul 2000, folosindu-se datele Federal Reserve Board’s CPI de la adresa http://minneapolisfed.org/research/data/ us/calc/. De exemplu, un milion de dolari din anul 1960 este echivalent cu 5, 8 milioane de dolari în anul 2000, iar un milion de dolari din anul 2004 este echivalent cu 0,91 milioane de dolari din anul 2000.

1949: http://www.el. cam.ac.uk/UNCCL/mişc/EDSAC99/statistics.html, http://www.davros.org/mişc/chronology.html.

1951: Richard E. Matick, *Computer Storage Sistems and Technology* (New York: John Wiley & Sons, 1977); http://inventors. about.com/library/ weekly/aaob2398.htm.

1955: Matick, *Computer Storage Sistems and Technology*, OECD, 1968, http://members. iinet.net.auz-dgreen/timeline.html.

1960: ftp: //rtfm. mit.edu/pub/usenet/alt. sys. pdp8/PDP-8 Frequently Asked Questions%28 postedeveryothermonth%29; http://www.dbit.com/-

greeng3/pdp1/pdp1.htm! INTRODUCTION.

1962: ftp: //rtfm. mit.edu/pub/usenet/alt. sys. pdp8/PDP-8 Frequently Asked Questions%28 postedeveryothermonth%29.

1964: Matick, *Computer Storage Sistems and Technology*, http://www.research. microsoft.com/users/ghell/craytalk; http://www.ddj.com/ documents/s = 1493/ddj0005 he/.

1965: Matick, *Computer Storage Sistems and Technology*, http://www.fourmilab.ch/documents/univac/config1108.htm l; http://www.frobenius.com/univac.htm.

1968: Data General.

1969, 1970: http://www.eetimes.com/special/specialissues/millennium/ milestones/whittier.html.

1974: Scientific Electronic Biological Computer Consulting (SCELBI). 1975 – 1996: publicitate din revista *Byte*.

1997 – 2000: publicitate din revista *PC Computing*.

2001: www.pricewatch.com (http://www.je-news.com/parse. cgi7 news/ pricewatch/raw/pw-010702).

2002: www.pricewatch.com (http://www.je-news.com/parse. cgi7 news/ pricewatch/raw/pw-020624).

2003: http://sharkyextreme.com/guides/WMPG/article php/10706 2227191.2.

2004: http://www.pricewatch.com (11/17/04).

2008: http://www.pricewatch.com (10/02/08) ($16.61).

15. Dataquest/intel şi Pathfinder Research:

Anul Dolari Logaritm natural

1968	1.00.000.000	O
1969	O.85.000.000	O.16252
1970	O.60.000.000	O.51083

1971	O.30.000.000	— 1.20397
1972	O.15.000.000	— 1.89712
1973	O.10.000.000	— 2.30259
1974	0.07.000.000	— 2.65926
1975	0.02800.000	— 3.57555
1976	0.01500.000	— 4.19971
1977	0.00800.000	— 4.82831
1978	0.00500.000	— 5.29832
1979	0.00200.000	— 6.21461
1980	0.00130.000	— 6.64539
1981	0,00082.000	— 7.10621
1982	0,00040.000	— 7.82405
1983	0,00032.000	— 8.04719
1984	0,00032.000	— 8.04719
1985	0,00015.000	— 8.80488
1986	0,00009.000	— 9.31570
1987	0,00008100	— 9.42106
1988	0,00006.000	— 9.72117
1989	0,00003500	—

				10.2602
				—
	1990		0,00002.000	10.8198
				—
	1991		0,00001700	10.9823
				—
	1992		0,00001.000	11.5129
	19		0,00000900	—
93				11.6183
	19			—
94			0,00000800	11.7361
	19			—
95			0,00000700	11.8696
	19			—
96			0,00000500	12.2061
	19			—
97			0,00000300	12.7169
	19			—
98			0,00000140	13.4790
	19			—
99			0,00000095	13.8668
	20			—
00			0,00000080	14.0387
	20			—
01			0,00000035	14.8653
	20			—
02			0,00000026	15.1626
	20			—
03			0,00000017	15.5875
	20			—
04			0,00000012	15.9358
	20		0,00000008	—
05		1		16.3288
	20		0,00000006	—
06		3		16.5801
	20		0,00000002	—
07		4		17.5452

	20	0,00000001	—
08	6		17.9507

16. Steve Cullen, InStat, septembrie 2008, www. instat.com.

An ul	*Mbiti*	*Bifi*
19 71	921.6	9.216E + 08
19 72	3788.8	3.789E + 09
19 73	8294.4	8.294E + 09
19 74	19865.6	1.987E + 10
19 75	42700.8	4.270E + 10
19 76	130662.4	1.307E + 11
19 77	276070.4	2.761 E + 11
19 78	663859.2	6.639E + 11
19 79	1438720.O	1.439E + 12
19 80	3172761.6	3.173E + 12
19 81	4512665.6	4.513E + 12
19 82	11520409.6	1.152E + 13
19 83	29648486.4	2.965E + 13
19 84	68418764.8	6.842E + 13
19 85	87518412.8	8.752E + 13
19	192407142.	1.924E +

86 4 14
19 255608422. 2.556E +
87 4 14
19 429404979. 4.294E +
88 2 14
19 631957094. 6.320E +
89 4 14
199 950593126.4 9.506
0 E + 14
199 1.547
1 1546590618 E + 15
199 2.846
2 2845638656 E + 15
199 4.178
3 4177959322 E + 15
199 7.511
4 7510805709 E + 15
’199 1.301
5 13010599936 E + 16
199 2.336
6 23359078007 E + 16
199 4.565
7 45653879161 E + 16
199 8.518
8 85176878105 E + 16
199 1.473
9 1.47327E + 11 E + 17
200 2.636
0 2.63636E + 11 E + 17
200 4.197
1 4.19672E + 11 E + 17
200 5.900
2 5.90009E + 11 E + 17
200 8.230
3 8.23015E + 11 E + 17
200 1.321
4 1.32133E + 12 E + 18

	200		1.995
5		1.9946E + 12	E + 18
	200		2.945
6		2.94507E + 12	E + 18
	200		5.628
7		5.62814E + 12	E + 18

17. „Historical Notes about the Cost of Hard Drive Storage Space”, http://www.littletechshoppe.com/ns1625/winchest.html; publicitate din revista *Byte*, 1977 – 1998; publicitate din revista *PC Computing*, 3/1999; *Understanding Computers*: *Memory*, *and Storage* (New York: Time Life, 1990); http://www.cedmagic.com/history/ibm-305 – ramac.html; John C. McCallum, „Disk Drive Prices (1955 – 2012)”, http://www.jemit.com/diskprice.htm; IBM, „Frequently Asked Questions”, http://www-03. ibm.com/ibm/ history/documents/pdf/fac.pdf; IBM, „IBM 355 Disk Storage Unit”, http://www-03. ibm.com/ibm/history/exhibits/storage/storage355.html; IBM, „IBM 3380 Direct Access Storage Device”, http://www.03-ibm.com/ibm/history/exhibits/storage/storage3380.html.

18. „Without Driver or Map, Vans Go from Italy to China”, *Sydney Morning Herald*, 29 octombrie 2010,http://www.smh.com.auztechnoiogy/technology-news/without-driver-or-map-vans-go-from-italy-to-china-20101029 – 176 ja.html.

19. KurzweilAI.net.

20. Material adaptat cu permisiunea lui Amiram Grinvald şi Rina Hildesheim, „VSDI: A New Era în Funcţional Imaging of Cortical Dynamics”, *Nature Reviews Neuroscience 5* (noiembrie 2004), pp. 874 – 85. Principalele instrumente pentru obţinerea de imagini ale creierului sunt expuse în această diagramă. Capacităţile lor sunt reprezentate de dreptunghiurile haşurate. Rezoluţia spaţială se referă la cea mai mică dimensiune care poate fi măsurată prin intermediul unei tehnici.

Rezoluţia temporală reprezintă durata imaginii. În cazul fiecărei tehnici, se fac concesii. De exemplu, EEG (electroencefalografia) care măsoară „undele cerebrale” (semnalele electrice ale neuronilor) poate măsura unde cerebrale foarte rapide (care se produc la intervale foarte apropiate), însă percepe numai semnalele din apropierea suprafeţei creierului.

În schimb, rezonanta magnetică funcţională FRMN, care foloseşte o maşină specială pentru, măsurarea fluxului de sânge la nivelul neuronilor (indicând activitatea acestora), poate să simtă mult mai profund în creier (şi la nivelul măduvei spinării) şi cu o rezoluţie superioară, de până la câteva zeci de microni (milionimi de metru), însă FRMN funcţionează foarte lent comparativ cu EEG.

Sunt tehnici neinvazive (nu necesită intervenţie chirurgicală sau administrarea de medicamente). MEG (magnetoencefalografia) este o altă tehnică neinvazivă. Detectează câmpurile magnetice generate de neuroni. MEG şi EEG pot să rezolve evenimente cu rezoluţie temporală de până la o millsecundă, dar mai bine decât FRMN, care poate cel mult să rezolve evenimente cu o rezoluţie de câteva sute de milisecunde. De asemenea, MEG identifică precis sursele din ariile primare senzoriale somatice şi motorii.

Imagistica optică acoperă aproape întreaga gamă de rezoluţii spaţiale şi temporale, dar este invazivă. Metoda indicatorilor sensibili la tensiuni electrice este cea mai potrivită metodă de măsurare a activităţii cerebrale, însă, la animale, se limitează la măsurătorile realizate imediat sub suprafaţa cortexului.

Cortexul expus este acoperit cu o incintă transparentă sigilată; după ce e imprimat cu un indicator sensibil la tensiuni electrice, este iluminat şi se imortalizează serii de fotografii cu o cameră de mare viteză. Alte tehnici optice folosite în laborator includ imagistica pe bază de ioni (de obicei de calciu sau sodiu) şi sistemele de imagine cu fluorescentă (imagistica

multifoton şi cea confocală).

Alte tehnici de laborator includ TEP (tomografia cu emisie de pozitroni, o tehnică imagistică de medicină nucleară care produce o imagine 3D), 2 DG (analiza postmortem cu 2-deoxiglucoză sau analiza tisulară), leziunile (presupune lezarea unor neuroni la animale şi observarea efectelor), clamparea pe bucăţi (pentru măsurarea curenţilor ionici pe membranele biologice) şi microscopia electronică (utilizarea unui fascicul de electroni pentru examinarea ţesuturilor la o scară foarte fină). Aceste tehnici pofti de asemenea integrate în imagistica optică.

21. Rezoluţia spaţială RMN în microni (pim), 1980 – 2012:

Anul Rezoluţia Citat URL

În microni

2012

125

„Characterization of Cerebral White Matter Properties Using http://dx. doi. Org/10.1089/ brain.2011.0071

2010 200

2010 250

1994 1.000

1989 1.700

1985 1.700

1983 1.700

1980 5.000

Quantitative Magnetic Resonance Imaging Stains”.

„Study of Brain Anatomy with High-Field MRI: Recent Progress”.

„High-Resolution Phased-Array MRI of the Human Brain at 7 Tesla: Iniţial Experience în Multiple Sclerosis Patients”.

„Mapping Human Brain Activity in Vivo”.

„Neuroimaging în Patients with Scizures of Probable Frontal Lobe Origin”.

„A Study of the Septum Pellucidum and Corpus Callosum în Schizophrenia with MR Imaging”.

„Clinical Efficiency of Nuclear Magnetic Resonance Imaging”.

„In Vivo NMR Imaging în Medicine: The Aberdeen Approach, Both Physical and Biological [and Discussion]” http://dx. doi.org/10.1089/

brain.2011.0071

http://dx. doi. Org/10.1016/j. mari.2010.02.007 http://dx. doi.org/10.1111

¡.1552 – 6569.2008.00338. X http://www.nebi. nim.

nih.gov/pâne/articles/

PMC1011409/

http://dx. doi.org/10.1111/

j.1528 – 157.1989.

M5470.X

http://dx. doi.org/10.1111/j.1600 – 0447.1985. th02634. x http://radiology. rsna.org/ content/146/1/123. short http://dx. doi.org/10.1098/

rsth.1980.0071

Anul Rezolup a Citatul URL Tehnica Note x-y (µm)

2011 4 „Focused Ion http://dx. doi. Focused ion beam/scan-Beam Miiling and org/10.3791/2588 ning electronmicroscope

Scanning Electron (FIB/SEM)

Microscopy of Brain Tissue”.

2011 4 „Volume http://dx. doi. Scanning electronmicro-Electron Microscopy org/10.1016/j. scopy (SEM)

for Neuronal Circuit comb.2011.10.022

Reconstruction”.

2011 4 „Volume http://dx. doi. Transmission electron

Electron Microscopy org/10.1016/j. microscopy (TEM)

for Neuronal Circuit comb.2011.10.022

Reconstruction”.

2004 13 „Serial Block-Face http://dx. doi. Serial

block-face scanning Rezultat citat în http://
Scanning Electron org/10.1371/joumal. electron microscopy faculty. ce. tamu.edu/
Microscopy to obio.0020329 (ZBF-SEM) choe/ftp/publications/
Reconstruct Threechoe. hpe08-preprint.
Dimensional Tissue pdf, asigurat de
Nanostructure" Yoonsuck Choe.
2004 20 „Wet SEM: A Novel http://dx. doi. „Wet" scanning electron
Method for Rapid org/10.1080/ microscopy (wet SEM)
Diagnosis of Brain 01913120490515603
Tumors".
1998 100 „A Depolarizing http://www.jneurosci. Scanning transmissio-Chloride Current org/content/18/17/6623. nelectron microscope
Contributes to full (STEM)
Chemoelectrical Transduction in Olfactory Senzory Neurons in Situ".
1994 2000 „Enhanced Optical http://joumals. lww. Enhanced optical imaging With a spaţial resolution of
Imaging of Rat Gliomas.com/neurosurgery/ the optical images below 20
and Tumor Margins" Abstract/1994/11.000/ microns 2/pixel (22).
Enhanced Optical lmagingof Rat
Gliomasand
Tumor.19.aspx
1983 3.000 „3D Imaging of X-Ray http://www.scipress. Projection microscopy Vezi Fig. 7 din articol.
Microscopy" org/e-library/sof2/
pdf70105. PDF
23. Rezoluţia spaţială în microni (pm) pentru tehnicile imagistice nedistructive la animale, 1985 – 2012:
A Rezol

nul	*uţia*		
2012	Resolution	0,07	
	Citation	Sebastián Beming şi colab., „Nanoscopy in a Living Mouse Brain”, *Science* 335, hr. 6068 (3 februarie 2012): 551. http://dx. doi. Org/10.1126/science.1215	
	URL	369	
	Technique	Stimulated emission depletion (STED) fluorescence nanoscopy	
	Notes	Highest resolution achieved in vivo so far	
2012	Resolution	0,25	
	Citation	Sebastián Beming şi colab., „Nanoscopy in a Living Mouse Brain”, *Science* 335, nr. 6068 (3 februarie 2012): 551. http://dx. doi. Org/10.1126/science.1215	
	URL	369	
	Technique	Confocal and multiphoton microscopy	
2004	Resolution	50	
	Citation	Amiram Grinvald şi Rina Hildesheim, „VSDI: A New Era în Funcţional Imaging of Cortical Dynamics”, *Nature Reviews Neuroscience 5* (noiembrie 2004): pp. 874	

– 85.

URL http://dx.doi.org/10.1038/nm1536

Technique Imaging based on voltage-sensitive dyes (VSDI)

Notes „VSDI has provided high-resolution maps, which correspond to cortical columns in which spiking occurs, and offer a spaţial resolution better than 50 pim”.

1996 Resolution 50

Citation Dov Malonek şi Amiram Grinvald. „Interactions between Electrical Activity and Cortical Microcirculation Revealed by Imaging Spectroscopy: Implications for Funcţional Brain Mapping”, *Science* 272, nr. 5261 (26 aprilie 1996): pp. 551 – 54.

URL

Technique

Notes

1995 Resolution

Citation

URL

Technique

Notes:

1985 Resolution Citation

URL

Technique http://dx.doi.org/10.1126/science.272.5261.551 Imaging spectroscopy „The study of spaţial relationships between individual cortical columns within a given brain area has become feasible with optical imaging based on intrinsic signals, at a spaţial resolution of about 50 pim”.

50

D.H. Turnbull and colab., „Ultrasound Backscatter Microscope Analysis of Early Mouse Embryonic Brain Development”, *Proceedings of the National Academy of Sciences* 92, nr. 6 (14 martie 1995): pp. 2239 – 43.

http://www.pnas. Org/content/92/6/2239. short Ultrasound backscatter microscopy „We demonstrate application of a real-time imaging method called ultrasound backscatter microscopy for visualizing mouse early embryonic neural tubes and hearts. This method was used to study live embryos in utero between 9.5 and 11.5 days of embryogenesis, with a spaţial resolution close to 50 nu”.

500

H.S. Orbach, L.B. Cohen şi A. Grinvald, „Optical Mapping of Electrical Activity în Rat Somatosensory and Visual Cortex”, *Journal of Neuroscience* 5, nr. 7 (01 iulie 1985): pp. 1886 – 95.http://www.jneurosci.org/content/

5/7/1886. short Optical methods

Capitolul 11. Obiecţii

1. Paul G. Allen şi Mark Greaves, „Paul Allen: The Singularity Isn’t Ne-ar”, *Technology Review*, 12 octombrie 2011,http://www.technologyreview.com/blog/ guest/27206/.

2. JTRS, „International Technology Roadmap for Semiconductors”, http://www.itrs.net/Links/20111TRS/Home2011.htm.

3. Ray Kurzweil, *The Singularity Is Ne-ar* (New York: Viking, 2005), Capitolul 2.

4. Nota de final 2 din Allen şi Greaves, „The Singularity Isn’t Ne-ar”, este formulată după cum urmează: „Începem să intrăm în gama de putere a computerelor de care am putea avea nevoie pentru susţinerea acestui gen de simulare masivă a creierului. Computerele din clasa Petaflop (cum ar.fi Blue Gene/P de la IBM, care a fost folosit pentru sistemul Watson) sunt disponibile acum în comerţ. Computerele din clasa Exaflop se află actualmente pe planşeta de proiectare. Aceste

sisteme ar putea, probabil, că dovedească acea capacitate computaţională brută necesară pentru simularea modelelor de excitaţie pentru toţi neuronii creierului, deşi actualmente o fac de multe ori mai lent decât s-ar întâmpla într-un creier real".

5. Kurzweil, *The Singularity Is* Wear. Capitolul 9, secţiunea intitulată „The Criticism from Software" (pp. 435 – 42).

6. Ibid., Capitolul 9.

7. Cu toate că nu este posibil să determinăm cu precizie conţinutul de informaţie al genomului, din cauza perechilor de baze repetate, aceasta [informaţia] este mult mai redusă decât datele totale necomprimate. Aici există două abordări pentru estimarea conţinutului de informaţie comprimată din genom, ambele demonstrând că o rată de 30 la 100 de milioane de bytes este suficient de ridicată.

1. În termeni de date necomprimate, există trei miliarde de trepte de ADN în codul genetic al omului, fiecare codificând câte 2 biţi (de vreme ce există patru posibilităţi pentru fiecare pereche de bază a ADN-ului). Prin urmare, genomul uman necomprimat este de aproximativ 800 de milioane de bytes. ADN-ul necodificat era numit odinioară „ADN junk", dar acum este clar că el joacă un rol important în expresia genelor. În orice caz, e codificat foarte ineficient. În primul rând, există redundanţe masive (de exemplu, secvenţa denumită „ALU* este repetată de sute de mii de ori), de care pot să profite algoritmii de compresie. Odată cu recenta explozie a băncilor de date genetice, există o doză de interes extraordinar de mare faţă de comprimarea datelor genetice. Cercetările recente asupra aplicării algoritmilor de date-standard pentru comprimare în cazul datelor genetice indică faptul că reducerea acestor date cu 90 la sută (la un bit de compresie perfectă) este fezabilă: Hisahiko Sato şi colab., „DNA Data Compression in the Post Genome Era", *Genome Informatics* 12 (2001): pp. 512 – 14, http://www.jsbi.org/joumal/GIW01/GIW01P130.pdf.

Prin urmare, putem comprima genomul la aproximativ 80 de milioane de bytes fără pierdere de informaţie (ceea ce înseamnă că putem reconstrui perfect întregul genom necomprimat de 800 de milioane de bytes).

Acum să ne gândim că peste 98 la sută din genom nu codifică proteine. Chiar şi după o comprimare-standard a datelor (care elimină redundanţele şi foloseşte o căutare de dicţionar pentru secvenţele obişnuite), conţinutul algoritmic al regiunilor necodificate pare să fie relativ redus, ceea ce înseamnă că e posibil să putem codifica un algoritm care ar îndeplini aceleaşi funcţii cu mai puţini biţi.

Pe de altă parte, de vreme ce ne aflăm abia la începutul procesului copierii genomului prin retroinginerie, nu putem face o estimare sigură a unei reduceri şi mai accentuate pe baza unui algoritm funcţional echivalent. Folosesc, prin urmare, o gamă de 30 până la 100 de milioane de bytes pentru informaţia comprimată din genom. Partea superioară a acestei game presupune numai comprimarea datelor, fără nicio simplificare algoritmică.

Doar o parte (deşi majoritară) a acestei informaţii caracterizează proiectul creierului.

2. O altă direcţie de raţionament este următoarea. Deşi genomul uman conţine în jur de 3 miliarde de baze, doar un mic procent, cum am menţionat mai înainte, codifică proteine. După estimările curente, există 26.000 de gene care codifică proteinele. Dacă presupunem că aceste gene conţin în medie 3.000 de baze de date utile, asta înseamnă doar aproximativ 78 de milioane de baze. O bază de ADN necesită numai 2 biţi, fapt ce se traduce prin aproximativ 20 de milioane de bytes (78 de milioane de baze împărţit la patru). În secvenţa genei de codificare a proteinei, fiecare „cuvânt” (codon) de trei baze ADN se traduce printr-un aminoacid. Prin urmare, există 43 (64) de posibile coduri de codon, fiecare constând în trei baze ADN. Pe de altă parte, există numai 20 de aminoacizi

folosiţi, plus un codon stop (aminoacid nul) din 64. Restul de 43 de coduri sunt folosite ca sinonime ale celor 21 de coduri utile. Deci sunt necesari numai 6 biţi pentru codificarea a 64 de combinaţii posibile şi doar aproximativ 4, 4 (log2 21) biţi sunt necesari pentru codificarea a 21 de posibilităţi, o economie de 1, 6 din 6 biţi (în jur de 27 la sută), ceea ce ne aduce la 15 milioane de bytes. În plus, aici este posibilă o comprimare-standard bazată pe repetarea secvenţelor, deşi pe această porţiune de codificare a proteinelor e posibilă o compresie mult mai redusă decât în cazul ADN-junk, posesor al unei masive redundanţe. Aşadar, ajungem la o cifră în jur de 12 milioane de bytes. Pe de altă parte, acum trebuie să mai adăugăm informaţie pentru porţiunea necodificată a ADN-ului care controlează expresia genelor. Deşi această porţiune de ADN constituie masa genomului, ea pare să aibă un nivel redus de informaţie şi e plină de redundanţe masive. Estimând că ar corespunde unui total de circa 12 milioane de bytes de coduri de proteine din ADN, ajungem din nou la aproximativ 24 de milioane de bytes. Din această perspectivă, o estimare de 30 până la 100 de milioane de bytes este moderat de ridicată.

8. Dharmendra S. Modha şi colab., „Cognitive Computing", *Communications of the ACM* 54, nr. 8 (2011): pp. 62 – 71, http://cacm. acm.org/ magazines/2011/8/114944-cognitive-computing/fulltext.

9. Kurzweil, *The Singularity Is Ne*-ar Capitolul 9, secţiunea intitulată „The Criticism from Ontology: Can a Computer Be Conscious?" (pp. 458 – 69).

10. Michael Denton, „Organism and Machine: The Flawed Analogy", din *Are* We *Spiritual Machines*? *Ray Kurzweil vs. the Critics of Strong Al* (Seattle: Discovery Institute, 2002).

11. Hans Moravec, *Mind Children* (Cambridge, MA: Harvard University Press, 1988).

Epilogul

1. „În U.S., Optimism about Future for Youth Reaches

All-Time Low", Gallup Politics, 2 mai 2011, http://www.gallup.com/poll/147350/optimism-future-youth-reaches-time-low. aspx.

2. James C. Riley, *Rising Life Expectancy*: *A Global History* (Cambridge: Cambridge University Press, 2001).

3. J. Bradford Delong, „Estimating World GDP, One Million B.C. – Present", 24 mai 1998, http://econ161.berkeley.edu/TCEH/1998 Draft/World GDP/Estimating World GDP.html şi http://futurist.typepad.com/myweblog/2007/07/economic-grawth.html. Vezi de asemenea Peter H. Diamandis şi Steven Kotler, *Abundance*: *The Future Is Better Than You Think* (New York: Free Press, 2012).

4. Martine Rothblatt, *Transgender to Transhuman* (publicată privat, 2011). Autoarea explică faptul că este posibil să apară o traiectorie la fel de rapidă a acceptării pentru „transumani", de exemplu, minţi nebiologice dar convingător de conştiente, după cum am discutat în Capitolul 9.

5. Următorul extras din *The Singularity Is Ne*-ar Capitolul 3 (pp. 133 – 35), de Ray Kurzweil (New York: Viking, 2005), discută limitele computaţiei bazate pe legile fizicii:

Limitele supreme ale computerelor sunt profund de înalte. Bazându-se pe lucrările profesorului Hans Bremermann de la Universitatea California din Berkeley şi ale teoreticianului nanotehnologiei Robert Freitas, profesorul Seth Lloyd de ia MIT a estimat capacitatea computaţională totală, potrivit legilor cunoscute ale fizicii, a unui computer cu greutatea de un kilogram şi ocupând volumul de un litru – cam de mărimea şi greutatea unui laptop mic –, ceea ce ei numeşte „laptopul suprem".

[Notă: Sâth Lloyd, „Ultimate Physical Limits to Computation", *Nature* 406 (2000): pp. 1047 – 54.

[Lucrările timpurii pe tema limitelor computaţiei au fost realizate de Hans J. Bremermann în anul 1962: Hans J. Bremermann, „Optimization Through Evolution and

Recombination", în M.C. Yovits, C.T. Jacobi, C.D. Goldstein, sub redacţia *Self-Organizing Sistems* (Washington, D.C.: Spartan Books, 1962), pp. 93 – 106.

[în anul 1984, Robert A. Freitas jr. a elaborat pe marginea lucrării lui Bremermann în Robert A. Freitas Jr., „Xenopsychology", *Analog* 104 (aprilie 1984): pp. 41 – 53, http://www.rfreitas.com/Astro/Xenopsychology.
htmăentience Quotient.]

Volumul potenţial de computaţie creşte odată cu energia disponibilă. Putem înţelege legătura dintre energie şi capacitatea computaţională după cum urmează. Energia dintr-o cantitate de materie este energia asociată cu fiecare atom (şi particulă subatomică). Aşadar, mai mulţi atomi, mai multă energie. După cum am mai spus, fiecare atom are potenţialul de a fi folosit pentru computaţie. Mai mulţi atomi, mai multă computaţie. Energia fiecărui atom sau particule creşte odată cu frecvenţa mişcării: mai multă mişcare, mai multă energie. Aceeaşi relaţie există şi pentru computaţia potenţială: cu cât este mai mare frecvenţa mişcării, cu atât e mai mare computaţia pe care o poate realiza fiecare componentă (care poate fi un atom). (Vedem acest lucru în tipurile contemporane: cu cât e mai mare frecvenţa tipului, cu atât este mai mare viteza de computaţie.)

Prin urmare, există o relaţie direct proporţională între energia unui obiect şi potenţialul său de a realiza computapa. Energia potenţială a unui kilogram de materie este foarte mare, aşa cum ştim din ecuapa lui Einstein $E = mc^2$. Viteza luminii reprezintă un număr foarte mare: aproximativ 1017 metru2/secundă2. Potenţialul de computare al materiei este guvernat de un număr, foarte, foarte mic, constanta lui Planck: 6, 6 * 1 (P joule-secundă (joule-ul este o unitate de măsură pentru energie). Aceasta este cea mai mică scară la care putem aplica energia pentru computape. Obţinem limita teoretică a unui obiect de a realiza computaţie împărţind energia totală (energia medie a fiecărui atom sau a fiecărei particule înmulţită cu

numărul de asemenea particule) la constanta lui Planck.

Lloyd arată că, pentru un kilogram de materie, capacitatea potenţială de computare este egală cu *pl* înmulţit cu energia şi împărţit la constanta lui Planck. Deoarece energia e un număr foarte mare, iar constanta lui Planck e atât de mică, această ecuaţie generează un număr uriaş: în jur de 5 * 1050 operaţii pe secundă.

[Notă: pl energia maximă (1017 kg metru2/secundă2) / (6, 6 * 10 34) joule-secundă) = – 5 * 1050 operaţii/secundă.]

Dacă asociem această valoare cu estimarea cea mai conservatoare a capacităţii creierelor umane (1019 cos şi 1010 oameni), ea reprezintă echivalentul a aproximativ 5 miliarde de mii de miliarde de civilizaţii umane.

[Notă: 5 x i050 cos este echivalentul a 5 * 1021 (5 miliarde de mii de miliarde) de civilizaţii umane (fiecare necesitând 1029 cos).]

Dacă folosim valoarea de 1016 cos, pe care o consider suficientă pentru imitarea funcţională a inteligenţei umane, laptopul suprem ar funcţiona cu echivalentul forţei cerebrale a 5 mii de miliarde de mii de miliarde de civilizaţii umane.

[Notă: Zece miliarde (1010) de oameni la 1016 cos fiecare înseamnă 1026 cos pentru civilizaţia umană. Deci 5 x10” cos este echivalentul a 5 x.1024 (5 mii de miliarde de mii de miliarde) de civilizaţii umane.]

Un asemenea laptop ar putea să realizeze echivalentul gândirii umane pe parcursul ultimelor zece mii de ani (respectiv zece miliarde de creiere omeneşti care operează timp de zece mii de ani) în a zece mia parte dintr-o nanosecundă.

[Notă: Această estimare emite premisa conservatoare că au existat zece miliarde de oameni în cursul ultimelor zece mii de ani, ceea ce evident că nu este cazul. Numărul real de oameni a crescut treptat în trecut pentru a ajunge la aproximativ 6, 1 miliarde în anul 2000. Există 3 x 107 secunde într-un an şi 3 x10” secunde în zece mii de ani.

Aşadar, folosind estimarea de 1026 cos pentru civilizaţia umană, gândirea umană pe parcursul ultimilor zece mii de ani este sigur echivalentă cu un mai mult de 3 x i037 calcule. Laptopul suprem realizează 5 x 1050 calcule într-o secundă. Aşadar, simularea a zece mii de ani de gândire a zece miliarde de oameni ar dura în jur de 10 – 13 secunde, ceea ce înseamnă a zece mia parte dintr-o nanosecundă.]

Din nou, câteva avertismente sunt bine venite. Convertirea masei laptopului nostru de un kilogram în energie reprezintă ceea ce se întâmplă practic în cazul unei explozii termonucleare: Desigur, nu vrem ca laptopul să explodeze, ci să rămână la dimensiunile lui cu un volum în jur de un litru. Aşadar, acest lucru va impune o ambalare atentă, ca să nu spunem mai mult. Analizând entropia maximă (gradele de libertate reprezentate de starea tuturor particulelor) unui asemenea dispozitiv, Lloyd arată că un laptop de felul acesta ar avea o capacitate teoretică a memoriei de 1031 biţi. Este dificil să-ţi imaginezi tehnologii care ar ajunge să atingă asemenea limite. Însă ne putem imagina foarte uşor tehnologii care se apropie rezonabil de mult. După cum arată proiectul Universităţii Oklahoma, am demonstrat deja capacitatea de a stoca cel puţin cincizeci de biţi de informaţie per atom (deşi numai pentru un mic număr de atomi, deocamdată). Stocarea a 1027 biţi de memorie în cei 1025 atomi dintr-un kilogram de materie ar trebui, prin urmare, să poată fi realizată în cele din urmă.

Însă, deoarece multe proprietăţi ale fiecărui atom pot fi exploatate pentru stocarea de informaţie – cum ar fi poziţia precisă, spinul şi starea cuantică a tuturor particulelor sale –, probabil că ne putem descurca într-un fel oarecare să depăşim 1027 biţi. Specialistul în neuroştiinţă Anders Sandberg estimează că potenţială capacitate de stocare a unui atom de hidrogen este de circa patru milioane de biţi. Aceste densităţi nu au fost încă demonstrate, aşa că vom folosi estimarea mai conservatoare.

[Notă: Anders Sandberg, „The Physics of the Information Processing Superobjects: Daily Life Among the Jupiter Brains”, *Journal of Evolution and Technology 5* (22 decembrie 1999),http://www.transhumanist.com/voLume5/Brains2.pdf.]

După cum am discutat mai înainte, 1042 calcule pe secundă se pot realiza fără producerea unei călduri semnificative. Folosind la maximum tehnicile reversibile de computare, utilizând formule care generează un nivel redus de erori şi permiţând o disipare rezonabilă a căldurii, ar trebui să ajungem undeva între 1042 şi 1050 calcule pe secundă.

Domeniul de concepţie dintre aceste două limite este complex. Examinarea subiectelor tehnice care intervin pe măsură ce avansăm de la 1042 la 1050 depăşeşte scopurile acestui capitol. Însă ar trebui să ţinem minte că rezolvarea nu va începe pornindu-se de la limita supremă de 1050, de la care să se scadă pe baza diverselor considerente practice. Dimpotrivă, tehnologia îşi va continua salturile, întotdeauna folosindu-şi capacitatea de ultimă oră pentru a progresa spre următorul nivel. Prin urmare, de îndată ce vom ajunge la o civilizaţie cu 1042 cos (la fiecare kilogram), oamenii de ştiinţă şi inginerii momentului respectiv îşi vor folosi inteligenţa în esenţă vast nonbiologică pentru a-şi da seama cum se poate ajunge la 1043, apoi la 1044 şi aşa mai departe. Eu, unul, mă aştept că ne vom apropia foarte mult de limitele supreme.

Chiar şi la 1042 cos, „computerul portabil suprem” de un kilogram ar putea să realizeze echivalentul gândirii umane din decursul ultimelor zece mii de ani (echivalat la zece miliarde de creiere umane timp de zece mii de ani) în zece microsecunde.

[Notă: Vezi nota de mai sus. 1042 cos este o valoare cu 10-e mai mică decât 1050 cos, aşadar a zece mia parte dintr-o nanosecundă devine 10 microsecunde.]

Dacă examinăm diagrama Creşterii Exponenţiale a Computaţiei (Capitolul 2), vedem că acest volum de

computaţie este estimat să fie disponibil contra sumei de o mie de dolari înaintea anului 2080.

CUPRINS

Mulţumiri 7

Introducere, 9

Capitolul 1. Experimente mintale despre lume 19

Capitolul 2. Experimentele mintale de gândire 30

Capitolul 3. Un model de neocortex: Teoria minţii bazate pe recunoaştere 38

Capitolul 4. Neocortexul biologic 76

Capitolul 5. Creierul vechi 92

Capitolul 6. Abilităţile transcendente 107

Capitolul 7. Neocortexul digital inspirat de biologie 118

Capitolul 8. Mintea ca un computer. 172

Capitolul 9. Experimente mintale despre minte 190

Capitolul 10. Legea accelerării veniturilor aplicată creierului 235

Capitolul 11. Obiecţii 252

Epilogul 262

Note 267